AD_____

Award Number: DAMD17-01-1-0720

TITLE: Structure-Function Relationships in Merlin, the Product of the NF2 Causal Gene

PRINCIPAL INVESTIGATOR: Zygmunt W. Derewenda, Ph.D.

CONTRACTING ORGANIZATION: University of Virginia
Charlottesville, VA 22904

REPORT DATE: October 2004

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188



| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | October 2004 | Annual (10 Sep 2003 - 9 Sep 2004) |

**4. TITLE AND SUBTITLE**
Structure-Function Relationships in Merlin, the Product of the NF2 Causal Gene

**5. FUNDING NUMBERS**
DAMD17-01-1-0720

**6. AUTHOR(S)**
Zygmunt W. Derewenda, Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Virginia
Charlottesville, VA 22904

E-Mail: Zsd4n@virginia.edu

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (Maximum 200 Words)**

Neurofibromatosis type 2 (NF2) is a debilitating, inherited disorder characterized by tumors of the brain and central nervous system. It is caused by the lack of a functional product of the *NF2* gene, known as merlin. Merlin serves as a linker protein between the cell membrane and the cytoskeleton. During the tenure of this grant, we used X-ray crystallography to determine the structure of the N-terminal FERM domain of merlin and we analyzed its functional implications. We also investigated structure-funtion relationships in syntenin, a scaffolding protein that binds to merlin. The structure of the PDZ tandem of syntenin was determined at 1.9 Å resolution, and we also determined the structure of the 2nd PDZ domain of syntenin to 0.73 Å, making it one of the highest resolution protein structures determined thus far. Furthermore, the structures of numerous complexes of syntenin with peptide ligands were determined to gain a better understanding of peptide recognition by PDZ domains. This lead to the formulation of the "combinatorial" model of peptide recognition, which helps explain non-canonical PDZ domain specificities. Biophysical and biochemical techniques were used to characterize the interaction of the PDZ domains with peptides derived from syntenin binding partners including merlin, syndecan, and the interleukin 5 receptor α-subunit.

**14. SUBJECT TERMS**
NF2, merlin, crystallography, syntenin, protein-protein interactions
Syndecan, interleukin 5 receptor α, PDZ domain.

**15. NUMBER OF PAGES**
106

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

## Table of Contents

# Introduction

Neurofibromatosis type 2 (NF2), characterized by tumors of the brain and central nervous system, is a debilitating, inherited disorder affecting 1 in 40,000 people (Evans et al., 2000). Genetic studies revealed that the disorder is caused by mutations in the *NF2* gene which codes for a protein known as merlin or schwannomin (Lutchman and Rouleau, 1996). Merlin is a member of the ERM (Ezrin, Radixin, and Moesin) family of proteins, but it is unique in that it acts as a tumor suppressor (Tsukita et al., 1997). Merlin exists in two different splice forms with different C-termini, has no catalytic activity, functions through interactions with other proteins and is regulated by phosphorylation (Rong et al., 2004). The aim of our Project was to provide insights into the atomic structure of merlin, and into the mechanisms of selected protein-protein interactions involving merlin, specifically the interaction with a regulatory protein RhoGDI which was thought to link merlin to the cytosolic GTPases such as RhoA. The study was subsequently extended to include another partner of merlin, a scaffolding protein syntenin.

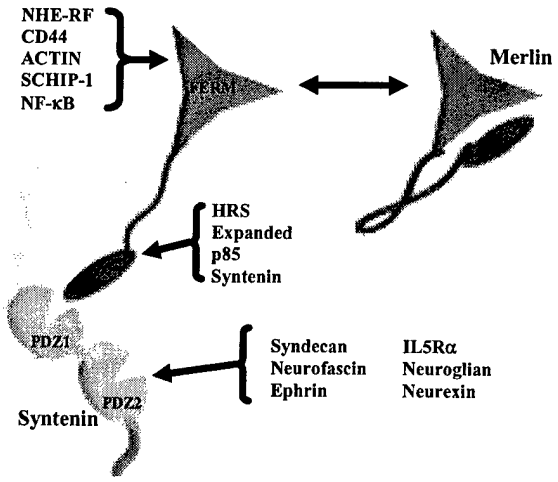



**Figure 1. A schematic representation of the closed and open conformations of merlin, and its interaction with syntenin. The diagram also lists other proteins implicated in interactions with either merlin or syntenin.**

# Body

## *1. Merlin - the product of the causal gene of NF2*

### *1.1. Determination of the N-terminal (FERM) domain structure.*

During the first year of this project, the structure of the N-terminal domain (FERM) domain, was determined by us using X-ray crystallography **(Kang et al., 2002)**. We expressed the FERM domain as an N-terminal GST fusion protein in *Escherichia coli* and purified by standard methods. The structure was solved by the molecular replacement method using the radixin FERM domain (Hamada et al., 2000) as the search model. The final model has an $R_{work}$ of 19.3% with an $R_{free}$ of 22.7 % and agrees well with standard protein geometry (other details can be found in the appended publication). The FERM domain is comprised of three subdomains, referred to as A, B, and C. Subdomain A is a mixed $\alpha$ / $\beta$ domain that resembles ubiquitin, subdomain B is a primarily helical domain with similarity to the acyl-CoA binding protein, and subdomain C is a mixed $\alpha/\beta$ domain that is similar to signaling domains such as PTB, PH and EVH1. The FERM domain of merlin is similar to the analogous domain of radixin and moesin (Pearson et al., 2000), albeit a number of differences are conspicuous.

Although the most severe cases of NF2 are caused by the complete or partial deletion of merlin, 20 missense mutations that cause NF2 are found in the FERM domain of merlin. The ways in which these mutations might affect the structure of merlin, and consequently cause NF2, are discussed in detail in the appended publication **(Kang et al., 2002)**. Overall, most of the mutations can be grouped into two categories: those that would disrupt the hydrophobic core of one of the FERM subdomains, and those that

would alter the 3D arrangement of the subdomains. The mutations that would disrupt the hydrophobic core of one of the subdomains most likely lead to an improperly folded protein that is targeted for degradation through the ubiquitin-proteasome pathway (Gautreau et al., 2002).

### *1.2 The interactions of the FERM domain with the C-terminal domain within merlin*

The N- and C-terminal domains of merlin can interact with each other, leading to closed and open conformations, with functional consequences. This head-to-tail interaction can be intramolecular or intermolecular, leading to 'closed' monomers or symmetric dimers, and similar interactions allow merlin to associate with other ERM proteins (Sherman et al., 1997b). As was outlined in our specific aims, we initiated studies of the interactions of these domains using isothermal calorimetry (ITC), a sensitive biophysical technique that had heretofore not been employed with any of the ERM proteins. The two isoforms of merlin differ in that marlin 2 terminates at exon 16 (C16) and merlin 1 excludes exon 16, but contains exon 17 (C17)(Haase et al., 1994). Merlin 1 is the major and most extensively researched isoform, and a predominance of the literature suggests that this is the only isoform capable of forming head to tail interactions.

To examine the interaction of the N- and C-termini of merlin, C16 and C17 oligopeptides were expressed as His-tagged proteins and purified using Ni-NTA and size exclusion columns. Prior to quantitating the interaction between these domains using ITC, other biochemical techniques were employed to verify that the constructs we had expressed were capable of interacting. Indeed, pull-down experiments where either the FERM domain or C17 was bound to its respective affinity column and the non-tagged

form of the other protein was passed through the column showed that the N- and C-terminal domains were able to specifically interact with each other. Moreover, when C17 and the FERM domain were mixed prior to size-exclusion chromatography, a peak corresponding to FERM-C17 heterodimer was observed. These initial experiments indicated that a more thorough examination of the FERM – C-terminal interaction was warranted.

In each ITC experiment, the FERM domain of merlin was placed in the cell of the ITC instrument and the C-terminal domain was added in small aliquots as the heat required to maintain the temperature of the system was monitored. All samples were extensively dialyzed into the same buffer before conducting the experiments to reduce systematic errors. A dissociation constant ($K_d$) of 96 nM and a stoichiometry of 1:1 was obtained for the interaction of the FERM domain and C17. Furthermore, our results indicate that there is also a weak 1:1 interaction between the FERM domain and C16, but an accurate dissociation constant could not be obtained. Representative data are presented in **Figure 2**.

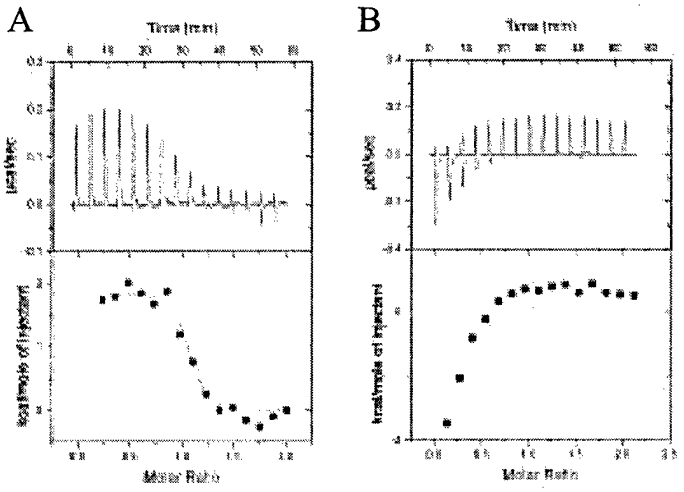



**Figure 2 Representative ITC data showing the association of the N- and C-terminal domains of merlin. A: The FERM domain and C17 (type 1 merlin) B: The FERM domain and C16 (type 2 merlin). The top panels show raw data while the bottom panels show the total amount of heat associated with each injection is plotted as a function of the molar ratio created in the cell by the injection.**

Our plan was to publish these data along with the crystal structure of the complex of the N- and C- terminal domains of merlin, or with the structure of the intact protein. Unfortunately, despite numerous efforts, we were not successful in obtaining

### *1.3. Interaction of Merlin with RhoGDI*

Data published by others strongly suggested that the FERM domain of merlin interacts with the Rho-specific nucleotide dissociation inhibitor (RhoGDI)(Maeda et al., 1999a). In fact it had been suggested that the interaction between ERM proteins and RhoGDI constitute the basis for the activation of Rho-GTPases by ERM proteins. Our previous work on RhoGDI (Longenecker et al., 1999; Longenecker et al., 2001), was originally one of the reasons why we were interested in pursuing the structural and functional studies of merlin. Unfortunately, contrary to earlier reports, we have been unable to see any evidence of a direct interaction of merlin's FERM domain with RhoGDI. When a mixture of RhoGDI and merlin were subjected to size exclusion chromatography, no peak corresponding to a complex was observed. Further, pull-down experiments also failed to detect any interaction. Finally, ITC experiments showed no heat associated with the mixing of these two proteins. Co-crystallization trials using the FERM domain and RhoGDI also failed to produce any crystals. We conclude that the earlier reports of ERM-RhoGDI interactions were based on experimental artifacts. It is noteworthy, that since this grant was awarded, no other reports of a RhoGDI – merlin interactions, or indeed ERM protein – RhoGDI interactions, have appeared in the literature.

## *2. Structural Studies of Syntenin*

With the consent of DOD and our reviewers, we shifted attention from RhoGDI to

syntenin, another partner of merlin (Jannatipour et al., 2001). This protein has also been shown to bind to a number of other proteins involved in cell proliferation and cellular regulation, making its association with a tumor suppressor, such as merlin, especially interesting. Syntenin is a cytosolic protein with two PDZ domains connected by a four residue linker. It was originally noted for its role in syndecan-mediated signaling. The PDZ tandem is preceded in syntenin by a 112 residue N-terminal domain of unknown function and followed by a short C-terminal domain (Grootjans et al., 1997). We determined the crystal structure of a number of variants of syntenin, alone and in the presence of peptide ligands, some that correspond to natural ligands and some that were designed to test the specificity of the two PDZ-domains. Our structural studies have provided significant insight into the mechanism of ligand recognition by PDZ domains of syntenin, and lead to a generalized proposal of the "combinatorial" model of peptide recognition by PDZ domains. This model accounts for many observations that were incongruous with the canonical model of PDZ peptide recognition.

All of the constructs of syntenin used in this study were generated using a clone of human syntenin obtained from ATCC (ATCC 72537). The gene was subcloned into the pGST-parallel 1 vector to insert the following fragments behind a GST affinity tag; full length (residues 1-298), PDZ1 (113-193), PDZ2 (197-273 and 197-270), Δ119 (120-298), PDZ2-C (197-298), and the tandem (residues 113-273). Expression and purification details are provided in the appended publications.

### *2.1. The structure of the PDZ tandem of syntenin*

The first study carried out by us in this area was that of the PDZ tandem (residues 113-273). This original crystal structure was solved using a three wavelength MAD

experiment at a synchrotron beamline X9B of the NSLS. The model was refined at 1.9 Å resolution to an $R_{work}$ of 17.7% and a $R_{free}$ of 23.6% **(Kang et al., 2003b)**. The atomic model revealed two PDZ tandems in the asymmetric unit arranged in a head-to tail fashion and related by a non-crystallographic two-fold symmetry axis. Like other domains from this superfamily, the syntenin PDZ modules show a typical fold with two opposing antiparallel β-sheets capped by two α-helices. Each domain has at least one β-strand that is partly contained in both sheets. There is some variation in the linker region between the two monomers, in part because of a very slight difference in the angles between the two PDZ domains in the different monomers. There is a quite extensive interface between the two monomers in the asymmetric unit.

Although the sequence identity between the two PDZ domains is modest (26%), the two domains are very similar to each other, with and rms deviation of only 1.2 Å for the Cα atoms. The main differences between the domains are the length of the β2-β3 loop, which is 4 residues longer in PDZ1, and the width of the peptide binding groove. A hydrogen bond between a α2 residue (Ser 170) and a backbone amide of β2 (Ser131) tethers the distal end of the peptide binding groove, making the helix to strand distance 1.8 Å shorter in PDZ1 than PDZ2.

The crystal structure suggested that syntenin has a defined supramodular architecture. This was further confirmed by the stability studies carried out in collaboration with Dr. Otlewski (University of Wroclaw, Poland). Solvent denaturation experiments showed that the isolated PDZ1 and PDZ2 have significantly different stabilities. The free energy of unfolding, $\Delta G_{un}$, for PDZ1 is -3.2 kcal/mole, while for PDZ2 it is -4.8 kcal/mole. The experimental unfolding of the tandem follows a

cooperative, two-state profile, with a $\Delta G_{un}$ of -4.1 kcal/mole. Further, the full-length protein unfolds in a highly cooperative manner, and shows significantly higher stability ($\Delta G_{un}$ of -6.4 kcal/mole) than any of the other constructs. It is therefore clear that the two PDZ domains interact within the protein and that the N- and C-terminal extensions also play a structural role **(Kang et al., 2003b)**.

To further investigate the structure of syntenin in solution, we recently used NMR spectroscopy (in collaboration with Dr. John Busweller, UVA). According to $^{15}N\{^{1}H\}$ NOEs and $^{15}N$ relaxation times, the PDZ tandem is monomeric in solution and the two domains tumble as a single unit with a rotational correlation time of 10 ns. The accurate arrangement of the PDZ domains in solution has been determined from residual dipolar couplings. While it is similar to the crystal structure, the domains are rotated by approximately -5°, 3°, and -23° about the x, y, and z axes. These different angels between domains found in the NMR and crystal structure suggest that the linker region is somewhat flexible and that the relative orientation of the two domains is not completely fixed, at least in the absence of the N- and C-terminal domains.

Furthermore, we analyzed NMR spectra of various syntenin fragments, including the full-length protein. The comparison of the $^{1}H$-$^{15}N$ HSQC spectra of full length syntenin to those of the PDZ tandem reveals an increased number of amide resonances with non-random coil chemical shifts. This indicates that there are fragments extraneous to the PDZ domains that are structured in the full-length protein.

The experimental work has just been completed and a manuscript was submitted for publication (**Cierpicki et al., 2004;** appended).

### *2.1. The structure of the isolated PDZ2 domain.*

In addition to the studies of the PDZ tandem of syntenin, we carried out extensive crystallization trials for the isolated PDZ domains. PDZ1 never crystallized, but we solved the structure of the isolated PDZ2 domain of syntenin. The original PDZ2 construct contained residues 197-273 and the crystals yielded an ultra-high resolution structure that was refined to 0.73 Å resolution with an R-factor of 7.5% and an $R_{free}$ of 8.7%, making it one of the highest resolution and most precisely refined protein structures determined to date **(Kang et al., 2004)**. Thus, the impact of this result extended well beyond the NF2 project.

The crystals of the isolated PDZ2 described above were not suitable for crystallographic studies of protein-peptide complexes because the peptide binding groove of one PDZ domain binds the C-terminus of a crystallographically related molecule in such a way, that the C-terminal phenylalanine of one PDZ domain occupies the $S_0$ pocket of another molecule, and the third residue of the construct (Met200) occupies the $S_{-2}$ site with hydrophobic interactions with Phe213. Thus, soaking in peptides or even co-crystallization is not feasible. To circumvent this problem, the last three residues of this construct were truncated resulting in a construct encompassing residues 197-270. This new construct proved to be suitable for co-crystallizations with peptides, as will be documented below.

### *2.2. The interaction of syntenin with peptides derived from its partner proteins.*

The apo-tandem crystal structure was not sufficient to predict which PDZ domain was responsible for binding which of the other numerous binding partners, including merlin.

In fact, the literature suggested that at least in some cases synergistic action by both PDZ domains is required for biological activity(Grootjans et al., 2000), although this could not be immediately rationalized by the structure. The canonical classification scheme for PDZ domains groups them into three classes based on the sequence of the ligand peptide. Class I domains have been defined as those that bind ligands with a C-terminal sequence containing a serine or threonine in $P_{-2}$ and a hydrophobic residue in $P_0$, so that the consensus motif is -S/T-X-Φ, where Φ represents a hydrophobic residues. Class II domains bind ligands with hydrophobic residues at $P_{-2}$ and $P_0$ (-Φ-X-Φ), while a negatively charged residue at $P_{-2}$ defines class III interactions (-D/E-X-Φ). Interestingly, although syntenin only has two PDZ domains, it has been shown to bind ligands of all three classes. Recently, several PDZ domains have also been found capable of binding more than one class of peptide. Thus, the shortcomings of this classification scheme are becoming readily apparent, and other groups suggested expanding the classification scheme to include novel classes (Bezprozvanny and Maximov, 2001). It is with these questions in mind that we conducted extensive structural and biophysical studies of the PDZ domains of syntenin in the presence of a number of peptides to determine the mechanism of peptide recognition.

Binding experiments were conducted to determine which PDZ domain of syntenin is responsible for binding which of the following three ligands: IL5Rα (class I), syndecan (class II) and merlin (class III) **(Kang et al., 2003b)**. Isothermal titration calorimetry (ITC) was used to determine the $K_d$ and stoichiometry of binding to syntenin. The syntenin constructs used in these first experiments were the full-length syntenin and the PDZ tandem. For merlin and IL5Rα, octapeptides and hexapeptides were evaluated. All

of the peptides bind to full-length syntenin and to PDZ tandem with dissociation constants ($K_d$) in the low μM range. For the merlin and syndecan peptides, a 1:1 binding stoichiometry was observed. Similarly, and 1:1 stoichiometry was found for the IL5Rα peptide and full-length syntenin, but a 2:1 ratio was found when this peptide was bound to the PDZ tandem. This not only suggests that this peptide is capable of binding to both PDZ domains, but also that the non-PDZ regions of syntenin posses a regulatory function. Furthermore, for the IL5Rα peptide, the length of the peptide did not have a significant affect on binding, but the merlin octapeptide binds an order of magnitude tighter than the hexapeptide, suggesting that the P-7 and P-8 residues play a role in determining binding specificity. Although experiments involving isolated PDZ2 were quite feasible, the poor solubility of the isolated PDZ1 precluded similar experiments and made it necessary to resort to fluorometric titrations using dansylated peptides, with much lower protein concentration required for measurements. These experiments were also conducted in collaboration with Dr. Otlewski. The IL5Rα peptide was found to interact with both PDZ domains, slightly more strongly with PDZ2. The merlin peptide was found to interact only with PDZ1 and the syndecan peptide interacted only with PDZ2.

We then used X-ray crystallography to determine the structure of the PDZ2 domain with two peptide ligands **(Kang et al., 2003a)**, and the structures of the tandem in complex with six peptides as described below **(Grembecka et al; in preparation)**.

In order to explain the degenerate specificity observed in the PDZ2 domain, we first solved crystal structures of PDZ2 in complexes with the IL5Rα and syndecan peptides. This study lead to a proposal of combinatorial peptide recognition **(Kang et al.,**

**2003a)**. Recently, we extended those studies to include the structures of the PDZ-tandem with peptides derived from two other partners of syntenin, i.e. neurexin 4 (TNEYYV)(Grootjans et al., 2000), and ephrin B (TNETKV)(Lin et al., 1999). We also designed peptides with non-natural sequences to probe the influence of particular residues on the peptide recognition mechanism. The peptide series TNEFYF, TNEFAF, and TNEAYF was used to determine the binding behavior when then peptide contained three aromatic residues and the influence of substitutions at the $P_{-1}$ and $P_{-2}$ position. These results are all new and are currently being prepared for publication. Basic crystallographic data for all these unpublished structures are presented in **Table I.** Because no other PDZ domain has had its structure determined with such an array of ligands, these studies will provide a wealth of information on the general mechanism of PDZ domain-ligand recognition and specificity, in addition to specific insight into the structure-function relationships of merlin.

**Table 1. Crystallographic data for the syntenin PDZ tandem – peptide complexes.**

| **Peptide** | **TNEFYF** | **TNEYYV** | **TNEYKV** | **TNEAYF** | **TNEFAF** | **ESYF** |
|---|---|---|---|---|---|---|
| Spacegroup | P41212 | P41212 | P41212 | P41212 | P41212 | P41212 |
| a=b (Å) | 72.207 | 72.134 | 72.341 | 71.918 | 72.192 | 72.091 |
| c (Å) | 126.046 | 126.221 | 125.676 | 127.342 | 125.887 | 126.828 |
| Resolution (Å) | 63.25 - 1.56 (1.60 - 1.56) | 8.00 - 2.25 (2.30 - 2.25) | 63.25 - 1.80 (1.85 - 1.80) | 62.02 - 1.80 (1.85 - 1.80) | 63.25 - 1.70 (1.74 - 1.70) | 63.25 - 1.70 (1.74 - 1.70) |
| Reflections (working) | 47197 | 15024 | 30277 | 28136 | 36073 | 36350 |
| Reflections (test) | 989 | 945 | 993 | 1197 | 1162 | 1169 |
| Completeness (%) | 100.0 | 99.6 | 98.7 | 92.4 | 99.6 | 99.9 |
| Rwork (%) | 17.1 (18.8) | 21.4 (25.0) | 18.1 (19.3) | 18.2 (20.3) | 17.3 (35.3) | 18.1 (23.5) |
| Rfree (%) | 20.1 (26.0) | 28.1 (33.0) | 22.9 (29.8) | 23.1 (29.6) | 22.6 (38.6) | 22.6 (31.0) |
| Number of waters | 381 | 200 | 281 | 325 | 370 | 395 |
| r.m.s. deviation from ideal geometry | | | | | | |
| Bonds (Å) | 0.017 | 0.011 | 0.016 | 0.011 | 0.012 | 0.018 |
| Angles (°) | 1.494 | 1.200 | 1.475 | 1.263 | 1.203 | 1.596 |
| Peptide in PDZ1 | Yes | No | No | Yes | No | Yes |
| Peptide in PDZ2 | Yes | Yes | Yes | Yes | Yes | Yes |

The structures of syntenin's PDZ tandem in complexes with peptides show many interesting and novel features that substantially alter our current understanding of these modules. With respect to the PDZ2 domains in all complexes, the terminal carboxylate of the peptide always accepts three hydrogen bonds from the amides of the "carboxylate binding loop" that precedes β2 ( Val209, Gly210, and Phe211 ) and the last side chain of the peptides fit into a hydrophobic $P_0$ pocket defined by Val209, Phe211, Phe213, and Leu258. However, although the classical model of PDZ domain peptide recognition requires that the $P_{-2}$ residue participate in specificity determination, the $P_{-2}$ Ser of IL5Rα (ETLEDSVF) and of the ESYF peptide does not directly interact with the PDZ domain and the peptide backbone is slightly displaced from α2 at the distal end of the binding pocket. Indeed, although these peptides are class I peptides, PDZ2 of syntenin does not even have a His at the beginning of α2 as is required by the classical model. In contrast to the missing interaction at $P_{-2}$, the side chains at $P_{-1}$ fit into a well defined hydrophobic $S_{-1}$ pocket formed by His208, Ile212, and Val222.

Although an interaction of the $P_{-1}$ residue with the PDZ domain is not included in the classical model, it is seen in a number of other structures determined in this study. In the PDZ2 structure with a syndecan peptide (TNEFYA), the $P_{-1}$ Tyr is situated in the $S_{-1}$ pocket with the aromatic ring of the tyrosine stacking against His208. Similar interactions are observed in other class II peptides, including ephrin B (TNEYYV), neurexin 4 (TNEYKV), and the hydrophobic peptides TNEFYF and TNEAYF. While the rest of the interactions for ephrin B and neurexin 4 are typical of class II interactions with PDZ domains, the TNEFYF and TNEAYF peptides do not interact with the $S_{-2}$ site.

Three of the peptides (TNEFYF, TNEAYF and ESYF) bound to both PDZ2 and

PDZ1. However, the binding of these peptides to PDZ1 is strikingly different from their binding to PDZ2, and deviates significantly from the canonical model. Although the C-terminal carboxylate, $P_0$ and $P_{-1}$ interactions are similar to the canonical interaction, the peptide then turns sharply, either crossing over β2 or exposing the peptide to solvent. An example of this type of non-canonical binding is seen in **Figure 3**.

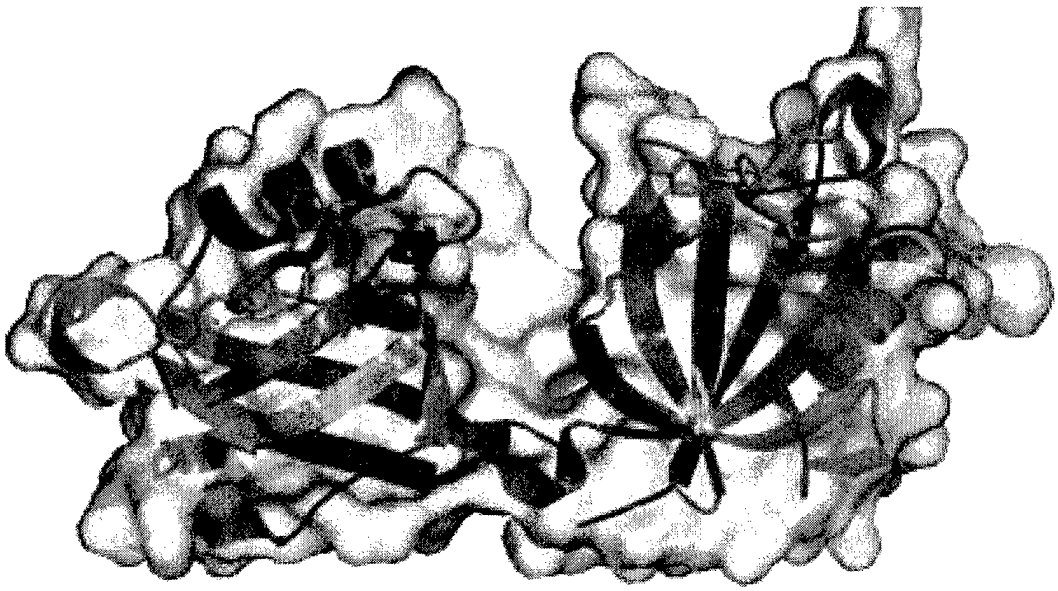


**Figure 3. The structure of the PDZ tandem of syntenin with an ESYF peptide. The syntenin structure is shown in a ribbon representation colored from blue (N-terminus) to red. PDZ1 is on the right and PDZ2 is on left. The peptide is represented by a coil for the main chain with all atoms for the side chains. Notice that although the peptide in PDZ2 follows the peptide binding groove, the peptide in PDZ1 only interacts in the C-terminal region of the peptide.**

The interaction of the syndecan (TNEFYA), neurexin (TNEYYV), ephrin B (TNEYKV) and the hydrophobic (TNEFYF) peptides with the PDZ tandem of syntenin was also investigated with NMR spectroscopy, to verify if the interactions see in the crystals represent those in solution. The binding of peptides was analyzed from a series of

$^{1}H$-$^{15}N$ HSQC spectra recorded for PDZ tandem titrated with increasing concentrations of peptides. All of the peptides were found to bind to both PDZ1 and PDZ2, albeit with significantly different changes in the chemical shifts for PDZ1 and PDZ2. The most noticeable changes are found within the PDZ2 domain, but there are significant perturbations of the chemical shifts for the TNEFYF and TNEYYV peptides. The weakest alterations resulted from the interaction of the TNEFAF peptide, which is consistent with the fact that this peptide was not found in the PDZ1 binding pocket in the X-ray structure. Figure 4 shows the crystal structure of syntenin PDZ tandem with the residue surfaces color coded to indicate the changes in the chemical shift upon peptide binding.

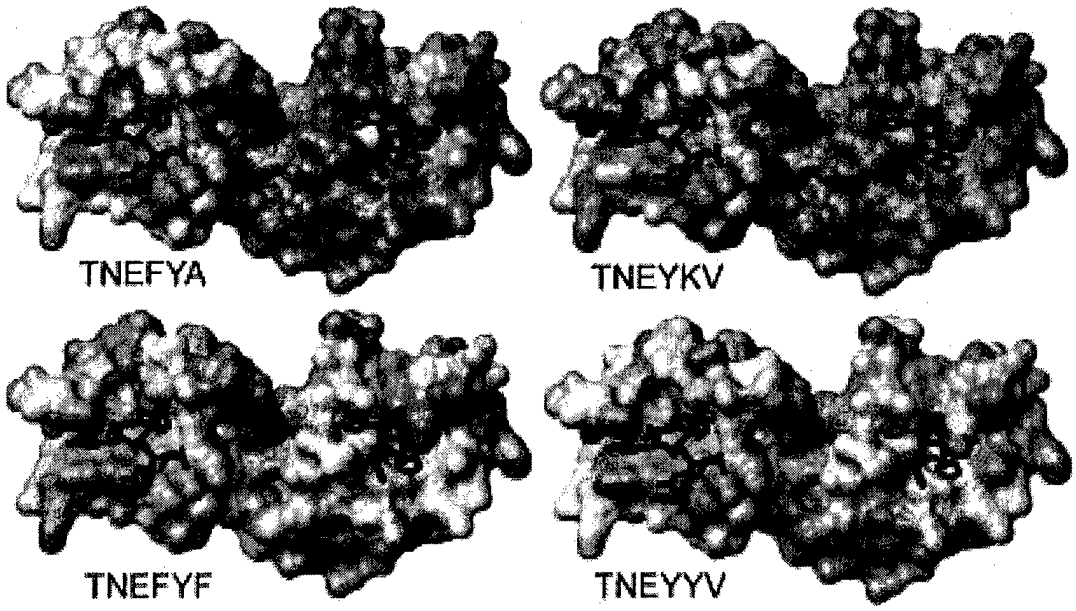



**Figure 4. Differences in NMR chemical shifts caused by peptide binding. The surface of the crystal structure of the PDZ tandem has been colored according to the magnitude of the chemical shift differences between the PDZ tandem and the PDZ tandem with peptide. Colors range from blue (no change) to red (largest change). The EFYA peptide is shown in each case for reference. In each case, PDZ1 is on the right and PDZ2 is on the left.**

NMR titration experiments were also used to assess the strength of interactions of the PDZ tandem with peptides. One of the advantages of this method is simultaneous quantification of interactions of the peptides with both domains. Binding constants have been obtained from the fitting of chemical shift changes as a function of a ligand concentration. For all peptides we observed fast exchange kinetics in agreement with moderate affinities obtained from ITC measurements for PDZ2. In addition, we were able to measure interactions with syntenin PDZ1 domain. Interestingly, the peptides derived from syndecan, neurexin and ephrin B bind to PDZ1 with at least 10 times weaker affinities than to PDZ2. The interaction of TNEYKV with PDZ1 is difficult to quantify, as it is weaker than 10 mM. Binding constants obtained using ITC and NMR titration experiments agree very well for the three hexapeptides. These data are presented in more detail in **Table II**.

**Table 2. Binding affinities of PDZ2 probed by ITC and NMR**

| | ITC (PDZ2) | NMR (PDZ Tandem) | |
|---|---|---|---|
| **Peptide** | **$K_d$ (μM)** | **$K_d$ PDZ1 (mM)** | **$K_d$ PDZ2 (μM)** |
| **TNEFYA** | 115 | 76 ± 1.6 | 160 ± 10 |
| **TNEFYF** | 98 | 1.00 ± 0.07 | 790 ± 30 |
| **TNEFAF** | ~600 | n.d. | n.d. |
| **TNEAYF** | ~1000 | n.d. | n.d. |
| **TNEYYV** | 96 | 0.74 ± 0.05 | 96 ± 9 |
| **TNEYKV** | >400 | >10 | 115 ± 90 |
| **EFYF** | 105 | n.d. | n.d. |
| **YYF** | 102 | n.d. | n.d. |

A large number of crystallographic experiments were conducted in an attempt to

characterize the interaction of merlin peptides in the active site of syntenin's PDZ domains. Both co-crystallization of the syntenin's PDZ tandem and octapeptide and hexapeptide and soaking these peptides into already formed crystals was attempted. Crystals generated using both of these techniques were used to collect diffraction data and the structures were determined. Unfortunately, no peptides were observed in the active site.

## Key Research Accomplishments

- X-ray crystallographic determination of the 1.8 Å resolution X-ray crystal structure of the N-terminal (FERM) domain of merlin, the product of the NF2 causal gene.
- X-ray crystallographic determination of the structure of the tandem PDZ domain of syntenin, alone and complexed with an assortment of peptides, including several corresponding to physiological binding partners of syntenin.
- Ultra-high resolution X-ray crystallographic determination of the PDZ2 domain of syntenin at 0.73 Å. This is currently the most precisely determined high resolution structure in the Protein Data Bank.
- X-ray crystallographic determination of the structure of the PDZ2 domain of syntenin complexed with an assortment of peptides, including several corresponding to physiological binding partners of syntenin.
- Calorimetric and biochemical analysis of the interaction of the PDZ domains of

syntenin with physiologically relevant peptides, including the identification of PDZ1 as the domain that binds merlin, PDZ2 as the domain that binds syndecan, and both as capable of binding IL5Rα.

- NMR determination of the structure of the PDZ tandem of syntenin alone and in complex with various peptides.
- Design and publication of the 'combinatorial' model of peptide recognition by PDZ domains.

## Reportable Outcomes

### *Publications:*

Beom Sik Kang, David Cooper, Yancho Devedjiev, Urszula Derewenda, Zygmunt Derewenda. *The structure of the FERM domain of merlin, the neurofibromatosis type 2 gene product.* Acta Crystallographica D Biological Crystallography. 2002. 58 (Pt 3):381-91.

Beom Sik Kang, David Cooper, Filip Jelen, Yancho Devedjiev, Urszula Derewenda, Zbigniew Dauter, Jacek Otlewski, Zygmunt S. Derewenda. *PDZ-tandem of human syntenin: crystal structure and functional properties.* Structure. 2003. 11:459-468.

Beom Sik Kang, David Cooper, Yancho Devedjiev, Urszula Derewenda, Zygmunt S. Derewenda. *Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm.* Structure. 2003. 11:845-853.

Beom Sik Kang, Yancho Devedjiev, Ulla Derewenda, Zygmunt Derewenda. *The PDZ2*

*Domain of Syntenin at Ultra-high Resolution: Bridging the Gap Between Macromolecular and Small Molecule Crystallography.* Journal of Molecular Biology. 2004. 338:483-493.

Tomasz Cierpicki, John H. Bushweller and Zygmunt S. Derewenda. *Probing the supramodular architecture of a multidomain protein: the solution structure of syntenin.* Submitted to Structure

**Two papers, dealing with the specificity of peptide recognition by syntenin, are currently in preparation.**

## *Presentations:*

**Oral Presentation:** David Cooper. *Crystal Structure of the FERM Domain of Merlin: The Neurofibromatosis 2 Tumor Suppressor Protein.* 29th Mid-Atlantic Crystallographic Workshop. Williamsburg, Virginia.

**Poster Presentation:** David Cooper, Beom Sik Kang, Peter Sheffield, Yancho Devedjiev, Zygmunt Derewenda. *Crystal Structure of the FERM Domain of Merlin, The Neurofibromatosis 2 Tumor Suppressor Protein.* American Crystallographic Association Annual Meeting. 2001, Los Angeles, California.

**Poster Presentation:** David R. Cooper, Beom Sik Kang, Yancho Devedjiev, Mary E. Lewis, Zbigniew Dauter, Zygmunt Derewenda. *The Tandem PDZ Domains of Syntenin.* 30th Mid-Atlantic Crystallographic Workshop. 2001, Frederick, Maryland.

**Poster Presentation:** David R. Cooper, Beom Sik Kang, Yancho Devedjiev, Ulla Derewenda, Mary E. Lewis, Zbigniew Dauter, Zygmunt Derewenda. *The Crystal*

*Structure of the PDZ Tandem of Syntenin.* American Crystallographic Association Annual Meeting. 2002, San Antonio, Texas. **Awarded Oxford Cryosystems Poster Prize.**

**Poster Presentation:** Jolanta Grembecka, Tomasz Cierpicki, Beom Sik Kang, Milton Brown, and Zygmunt Derewenda. T*owards Rational Design of Selective Ligands for Syntenin PDZ Domains* Structure-based drug design conference. 2003, Boston, MA.

## *Protein Data Bank Depositions:*

1H4R - Crystal Structure Of The FERM Domain Of Merlin, The Neurofibromatosis 2 Tumor Suppressor Protein.

1N99 - Crystal Structure Of The PDZ Tandem Of Human Syntenin.

1NTE - Crystal Structure Analysis of The Second PDZ Domain Of Syntenin.

1OBX - Crystal Structure Of The Complex Of PDZ2 Of Syntenin With An Interleukin 5 Receptor Peptide.

1OBY - Crystal Structure Of The Complex Of PDZ2 Of Syntenin With A Syndecan-4 Peptide.

1OBZ - Crystal Structure Of The Complex Of The PDZ Tandem Of Syntenin With An Interleukin 5 Receptor Peptide.

1R6J – Ultra-high Resolution Crystal Structure Of Syntenin PDZ2.

1V1T - Crystal structure of the PDZ tandem of human syntenin with TNEYKV peptide.

**The following structures will be deposited in the protein data bank.**

- PDZ tandem of syntenin with ESYF peptide

- PDZ tandem of syntenin with TNEAYF peptide
- PDZ tandem of syntenin with TNEFYF peptide
- PDZ tandem of syntenin with TNEFYF peptide
- PDZ tandem of syntenin with TNEYYV peptide

## *All personnel involved at different stages of the project*

**Principle Investigator**

Zygmunt Derewenda

**Research Faculty**

- Urszula Derewenda
- Yantcho Devedjiev

**Research Associates**

- Tomasz Cierpicki
- David Cooper
- Jolanta Grembecka
- Beom Sik Kang

**Technical Assistants**

- Holly Barton
- Neelima Choudhary
- Mary Lewis
- Natalya Oleknovich

## *Conclusions*

## *Determination of the N-terminal (FERM) domain structure*

The structure of the FERM domain of merlin yields significant insight into the function

of merlin. Although the structures of several other FERM domains have now been solved (Hamada et al., 2000; Pearson et al., 2000), the merlin structure is critical for the understanding neurofibromatosis type 2, since merlin is the only ERM protein that has a tumor suppressor function (Gusella et al., 1999). With that in mind, the structure of the FERM domain was analyzed to see how it relates to NF2. A large number of missense mutations are speculated to alter the three dimensional arrangement of the subdomains, suggesting that the overall tertiary structure of the FERM domain is crucial for the proper function of merlin. The majority of the missense mutations that affect the tertiary fold of the FERM domain do so by altering the interface between the first and second subdomains. This fact by itself might be overlooked, except for the fact that this interface has several other unique characteristics. First, it is the only region of the FERM domain of merlin that has a dramatically different electrostatic potential from other published FERM domains. Secondly, it is flanked by clusters of residues that are conserved among the other FERM domains, yet different in merlin. These patches of residues create surface epitopes that are unique to merlin among the FERM proteins and probably affect the way merlin interacts with effectors. Taken together, this implicates the face of the FERM domain that is shared by the first and second subdomains of merlin as critical for the tumor suppressing function of merlin. A more thorough discussion of the conclusions of the structural determination of the FERM domain of merlin is detailed in the Appendix (Kang et al., 2002).

The solution of the crystal structure of merlin's FERM domain is the most important accomplishment of this project. Merlin is the sole key molecule in NF2 pathogenesis, and the structure of this molecule puts the structural biology relevant to the

disease on a strong footing.

The interaction of the N- and C-terminus of merlin is well documented, but the prevalent view in the literature is that only merlin 1, with a C-terminal tail encoded by exon 17, is able to interact with the FERM domain. The significance of this is enhanced by the fact that only merlin 1 shows a tumor suppressor function (Pearson et al., 2000; Sherman et al., 1997a). Our experiments certainly confirm that the N- and C-termini of merlin 1 do interact with a $K_d$ of 96 nM , but also detect a weak, but present, interaction of the C-termini of merlin 2. It is possible that we were able detect this interaction specifically due to the sensitive nature of isothermal titration calorimetry.

## *The interaction of merlin with RhoGDI*

At the time when our proposal was being prepared, there was a significant level of excitement in the community associated with the reported observations suggesting that some ERM proteins, including merlin, interact with RhoGDI (Maeda et al., 1999b; Takahashi et al., 1997). Given our interest in RhoGDI and RhoA-mediated signaling pathways (Longenecker et al., 1999; Wei et al., 1997), we were keen to pursue this avenue of research and we had the means to do it. However, all our assays returned negative data. It is also interesting that although there has been one report of co-crystallization of RhoGDI with the FERM domain of radixin, a homologue of merlin, (Hamada et al., 2001), this structure was never published. This strongly suggests that the crystals did not contain the complex, as the original publication stated. Under the circumstances, we discontinued this research and we focused on another partner of merlin - syntenin.

## *The structure and function of syntenin*

Although many PDZ-containing proteins have multiple PDZ domains, our structure of the tandem of PDZ domains of syntenin was the first reported crystal structure to contain more than one PDZ domain within a contiguous polypeptide chain **(Kang et al., 2003b)**. Each PDZ domain in the tandem appeared capable of binding peptides, as these domains deviated very little from a classical PDZ fold. One exception to this is the insertion of a basic residue after the initial glycine of the signature peptide-binding GLGF loop. The crystal structure, along with stability studies, strongly suggested that syntenin has a supramodular architecture, and that the mutual disposition of the two PDZ domains is fixed. This was a novel concept at the time, but soon other examples were documented (Long et al., 2003). Since crystal structures are occasionally biased due to the crystal packing forces, we pursued this question with a study of the PDZ-tandem and intact syntenin using a relatively novel NMR technique based on the measurements of residual dipolar couplings (RDCs). We showed (manuscript submitted) indeed syntenin has a stable supramodular structure, but that it is not identical to the one in the crystal structure, because of the crystal packing.

The studies of syntenin-peptide complexes revealed several surprises, and resulted in observations whose significance extends beyond the structural biology of NF2. We discovered the basis of degenerate specificity in the PDZ2 domain of syntenin and we revised the current theory regarding the molecular recognition of peptides by the PDZ domains. This work was featured on the cover of the July 2003 issue of the Cell Press journal - STRUCTURE. Moreover, the crystal structure of the isolated PDZ2 domain set a new record for the highest precision in crystallographic analysis. The ultra-high

resolution structure of PDZ2 proved not only to be informative for the study of the mechanism of PDZ peptide recognition, but due to the unprecedented quality of the data, it set a new standard for protein crystallography.

Further, in the last year, we solved several structures of complexes of the PDZ-tandem with various peptides, probing the nature of the differences between the two domains. We characterized a new, non-canonical mode of interaction, unique to the PDZ1 domain. Again, these results significantly extend our understanding of the structural biology of the PDZ domains.

## References


Bezprozvanny, I., and Maximov, A. (2001). Classification of PDZ domains. FEBS Lett *509*, 457-462.

Evans, D. G., Sainio, M., and Baser, M. E. (2000). Neurofibromatosis type 2. J Med Genet *37*, 897-904.

Gautreau, A., Louvard, D., and Arpin, M. (2002). ERM proteins and NF2 tumor suppressor: the Yin and Yang of cortical actin organization and cell growth signaling. Current Opinion in Cell Biology *14*, 104-109.

Grootjans, J. J., Reekmans, G., Ceulemans, H., and David, G. (2000). Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin. Journal of Biological Chemistry *275*, 19933-19941.

Grootjans, J. J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Durr, J., and David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. Proceedings of the National Academy of Sciences of the United States of America *94*, 13683-13688.

Gusella, J. F., Ramesh, V., MacCollin, M., and Jacoby, L. B. (1999). Merlin: the neurofibromatosis 2 tumor suppressor. Biochimica et Biophysica Acta *1423*, M29-36.

Haase, V. H., Trofatter, J. A., MacCollin, M., Tarttelin, E., Gusella, J. F., and Ramesh, V. (1994). The murine NF2 homologue encodes a highly conserved merlin protein with alternative forms. Human Molecular Genetics *3*, 407-411.

Hamada, K., Seto, A., Shimizu, T., Matsui, T., Takai, Y., Tsukita, S., and Hakoshima, T. (2001). Crystallization and preliminary crystallographic studies of RhoGDI in complex with the radixin FERM domain. Acta Crystallogr D Biol Crystallogr *57*, 889-890.

Hamada, K., Shimizu, T., Matsui, T., Tsukita, S., and Hakoshima, T. (2000). Structural basis of the membrane-targeting and unmasking mechanisms of the radixin FERM domain. EMBO Journal *19*, 4449-4462.

Jannatipour, M., Dion, P., Khan, S., Jindal, H., Fan, X., Laganiere, J., Chishti, A. H., and Rouleau, G. A. (2001). Schwannomin isoform-1 interacts with syntenin via PDZ

domains. Journal of Biological Chemistry *276*, 33093-33100.

Kang, B. S., Cooper, D. R., Devedjiev, Y., Derewenda, U., and Derewenda, Z. S. (2002). The structure of the FERM domain of merlin, the neurofibromatosis type 2 gene product. Acta Crystallographica Section D-Biological Crystallography *58*, 381-391.

Kang, B. S., Cooper, D. R., Devedjiev, Y., Derewenda, U., Derewenda, Z. S., Jelen, F., Dauter, Z., and Otlewski, J. (2003a). Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. Structure (Camb) *11*, 845-853.

Kang, B. S., Cooper, D. R., Jelen, F., Devedjiev, Y., Derewenda, U., Dauter, Z., Otlewski, J., and Derewenda, Z. S. (2003b). PDZ tandem of human syntenin: crystal structure and functional properties. Structure (Camb) *11*, 459-468.

Kang, B. S., Devedjiev, Y., Derewenda, U., and Derewenda, Z. S. (2004). The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography. J Mol Biol *338*, 483-493.

Lin, D., Gish, G. D., Songyang, Z., and Pawson, T. (1999). The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif. Journal of Biological Chemistry *274*, 3726-3733.

Long, J. F., Tochio, H., Wang, P., Fan, J. S., Sala, C., Niethammer, M., Sheng, M., and Zhang, M. (2003). Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. J Mol Biol *327*, 203-214.

Longenecker, K., Read, P., Derewenda, U., Dauter, Z., Liu, X., Garrard, S., Walker, L., Somlyo, A. V., Nakamoto, R. K., Somlyo, A. P., and Derewenda, Z. S. (1999). How RhoGDI binds Rho. Acta Crystallographica Section D-Biological Crystallography *55*, 1503-1515.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J., and Derewenda, Z. S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. Acta Crystallographica Section D-Biological Crystallography *57*, 679-688.

Lutchman, M., and Rouleau, G. A. (1996). Neurofibromatosis type 2: a new mechanism

of tumor suppression. Trends Neurosci *19*, 373-377.

Maeda, M., Matsui, T., Imamura, M., and Tsukita, S. (1999a). Expression level, subcellular distribution and rho-GDI binding affinity of merlin in comparison with Ezrin/Radixin/Moesin proteins. Oncogene *18*, 4788-4797.

Maeda, M., Matsui, T., Imamura, M., and Tsukita, S. (1999b). Expression level, subcellular distribution and rho-GDI binding affinity of merlin in comparison with Ezrin/Radixin/Moesin proteins. Oncogene *18*, 4788-4797.

Pearson, M. A., Reczek, D., Bretscher, A., and Karplus, P. A. (2000). Structure of the ERM protein moesin reveals the FERM domain fold masked by an extended actin binding tail domain. Cell *101*, 259-270.

Rong, R., Surace, E. I., Haipek, C. A., Gutmann, D. H., and Ye, K. (2004). Serine 518 phosphorylation modulates merlin intramolecular association and binding to critical effectors important for NF2 growth suppression. Oncogene.

Sherman, L., Xu, H. M., Geist, R. T., Saporito-Irwin, S., Howells, N., Ponta, H., Herrlich, P., and Gutmann, D. H. (1997a). Interdomain binding mediates tumor growth suppression by the NF2 gene product. Oncogene *15*, 2505-2509.

Sherman, L., Xu, H. M., Geist, R. T., Saporito-Irwin, S., Howells, N., Ponta, H., Herrlich, P., and Gutmann, D. H. (1997b). Interdomain binding mediates tumor growth suppression by the NF2 gene product. Oncogene *15*, 2505-2509.

Takahashi, K., Sasaki, T., Mammoto, A., Takaishi, K., Kameyama, T., Tsukita, S., and Takai, Y. (1997). Direct interaction of the Rho GDP dissociation inhibitor with ezrin/radixin/moesin initiates the activation of the Rho small G protein. J Biol Chem *272*, 23371-23375.

Tsukita, S., Yonemura, S., and Tsukita, S. (1997). ERM proteins: head-to-tail regulation of actin-plasma membrane interaction. Trends in Biochemical Sciences *22*, 53-58.

Wei, Y., Zhang, Y., Derewenda, U., Liu, X., Minor, W., Nakomoto, R. K., Somlyo, A. V., Somlyo, A. P., and Derewenda, Z. S. (1997). Crystal structure of RhoA-GDP and its functional implications. Nature Structural Biology *4*, 699-702.

Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# The structure of the FERM domain of merlin, the neurofibromatosis type 2 gene product

**Beom Sik Kang,† David R. Cooper,† Yancho Devedjiev, Urszula Derewenda and Zygmunt S. Derewenda***

Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908-0736, USA

† These authors contributed equally to the project.

Correspondence e-mail: zsd4n@virginia.edu

Received 23 August 2001
Accepted 10 December 2001

**PDB Reference:** merlin FERM domain, 1h4r, r1h4rsf.

Neurofibromatosis type 2 is an autosomal dominant disorder characterized by central nervous system tumors. The cause of the disease has been traced to mutations in the gene coding for a protein that is alternately called merlin or schwannomin and is a member of the ERM family (ezrin, radixin and moesin). The ERM proteins link the cytoskeleton to the cell membrane either directly through integral membrane proteins or indirectly through membrane-associated proteins. In this paper, the expression, purification, crystallization and crystal structure of the N-terminal domain of merlin are described. The crystals exhibit the symmetry of space group $P2_12_12_1$, with two molecules in the asymmetric unit. The recorded diffraction pattern extends to 1.8 Å resolution. The structure was solved by the molecular-replacement method and the model was refined to a conventional $R$ value of 19.3% ($R_{\rm free}$ = 22.7%). The N-terminal domain of merlin closely resembles those described for the corresponding domains in moesin and radixin and exhibits a cloverleaf architecture with three distinct subdomains. The structure allows a better rationalization of the impact of selected disease-causing mutations on the integrity of the protein.

## 1. Introduction

Neurofibromatosis type 2 (NF2), first described in 1822 by the Scottish surgeon Wishart, is an often devastating autosomal dominant disorder affecting one in every 40 000–90 000 potential births, depending on geographic factors (Evans *et al.*, 1992, 2000; Gutmann, 2001; Martuza & Eldridge, 1988). Until about 1985, NF2 was often linked with neurofibromatosis type 1, also a dominant inherited disorder, and the two were collectively referred to as von Recklinhausen disease. Individuals affected by NF2 develop central nervous system tumors such as Schwann cell tumors of the eighth cranial nerve (bilateral vestibular schwannomas), meningiomas and ependymomas, which although classified as cancers are typically slow-growing and non-malignant. The clinical symptoms vary profoundly from a mild to a very severe phenotype, with diagnostic prevalence of the disease significantly lower than birth incidence (Evans *et al.*, 2000; Gutmann, 2001).

Neurofibromatosis type 2 is associated with a homozygous inactivation of the NF2 gene. Located within 17 exons in the long arm of chromosome 22, this gene encodes a 595-residue protein denoted as schwannomin or merlin (Rouleau *et al.*, 1993; Trofatter *et al.*, 1993). Alternative splicing of exon 16 results in the presence of another isoform, which differs only in the C-terminal 11 residues, with important functional consequences (Sherman *et al.*, 1997). There is convincing evidence that mutations inactivating some or all of the biological functions of merlin, which acts as a tumor

suppressor protein, are the causal factor behind the etiology of NF2. For example, in schwannoma, overexpression of wild-type NF2 gene but not of a mutant leads to growth suppression, impaired cell motility, adhesion and spreading (Gutmann *et al.*, 1998, 1999; Sherman *et al.*, 1997). Furthermore, mice with targeted mutations in the NF2 gene develop malignant tumors (McClatchey *et al.*, 1998).

Merlin is a member of a larger group of proteins, which includes protein 4.1, talin and three closely homologous proteins known collectively as ERM, *i.e.* ezrin, radixin and moesin (Mangeat *et al.*, 1999; Tsukita *et al.*, 1994, 1997; Tsukita & Yonemura, 1997). The ERM proteins have no known catalytic function, but are believed to participate in signaling phenomena by providing a link between the actin cytoskeleton and the membrane (Tsukita *et al.*, 1994). Like other ERM proteins, merlin contains three domains: the N-terminal domain (also denoted as the FERM domain) comprising approximately the first 300 residues, a central coiled-coil fragment and a C-terminal polypeptide containing the last 120 residues. The C-terminal polypeptide of merlin is unique among the ERM family members in that it does not contain an actin-binding motif (Mangeat *et al.*, 1999; Turunen *et al.*, 1998). The molecular physiology of merlin and of the ERM proteins in general involves intermolecular or intramolecular head-to-tail interaction between the FERM domain and the C-terminal polypeptide (Meng *et al.*, 2000; Nguyen *et al.*, 2001; Sherman *et al.*, 1997; Tsukita *et al.*, 1997). The FERM domain of merlin has been implicated in intermolecular interactions with such proteins as CD44 (Herrlich *et al.*, 2000), EBP50 (NHE-RF; Murthy *et al.*, 1998), SCHIP-1 (Goutebroze *et al.*, 2000), HRS (Scoles *et al.*, 2000), $\beta$1-integrin (Obremski *et al.*, 1998) and RhoGDI (Maeda *et al.*, 1999). Whether or not all of these interactions are physiologically relevant remains to be validated, as are the specific signaling pathways relevant to merlin. However, the regulated association of the FERM domain of merlin with the C-terminal polypeptide (also denoted C-ERMAD) mediates tumor-growth suppression in normal cells (Sherman *et al.*, 1997). Under normal conditions the association between the two domains is regulated by phosphorylation of the C-terminal polypeptide, although it is not clear what induces this process.

Recently, the molecular architecture of the ERM proteins has become better understood owing to X-ray diffraction analyses of the FERM domains of radixin and moesin. The moesin domain structure was solved at 1.9 Å resolution in complex with its partner C-terminal polypeptide, but with the intervening coiled-coil fragment removed by recombinant methods (Pearson *et al.*, 2000), and was also studied independently in a form which includes an extension into the coiled-coil region at 2.7 Å resolution (Edwards & Keep, 2001). The radixin FERM domain was solved with and without bound inositol-(1,4,5)-triphosphate (IP3) at 2.8 and 2.9 Å resolution, respectively (Hamada *et al.*, 2000). In addition, a more distantly related domain from protein 4.1 was also solved by X-ray diffraction at 2.8 Å resolution (Han *et al.*, 2000). These studies revealed that the FERM domains are structurally very similar, with a cloverleaf-like architecture consisting of three distinct subdomains. The N-terminal subdomain has a ubiquitin-like fold and is followed by a subdomain resembling an acyl-CoA binding protein and a third subdomain reminiscent of a phosphotyrosine-binding domain (PTB) or pleckstrin homology domain (PH). In the structure of the moesin intramolecular complex (Pearson *et al.*, 2000), the C-terminal polypeptide adopts an extended meandering conformation, which suggests that without its FERM partner it is unable to form a stable tertiary fold.

In spite of significant progress in the studies of ERM proteins, the structure of merlin, the specific molecule associated with NF2, has not been described. It is important to stress that there are critical functional differences between merlin and its homologs and that only merlin mutations are associated with the neurofibromatosis phenotype. Efforts to design therapeutic agents able to interact with the FERM domain of merlin in a way that could relieve NF2 symptoms would certainly benefit from an accurate knowledge of the molecular structure of merlin itself. Here, we report the structure of the human merlin FERM domain (residues 1–313) at 1.8 Å resolution. The structure reveals the expected conserved cloverleaf architecture of the FERM domain and provides an additional rationale for the pathological effects of the known NF2-associated missense mutations. It also suggests regions of the protein that are critical for the interactions with effectors and/or activators of merlin.

## 2. Materials and methods

### 2.1. Construction of FERM domain expression plasmids

A merlin clone was purchased from American Tissue and Culture Collection (ATCC 106908). It contained the nucleotide sequence corresponding to the merlin N-terminal 341 amino acids. To express this sequence using the Gateway gene-expression system (Life Technologies) and to introduce the recombinant TEV protease (rTEV) cleavage site between the glutathione S-transferase (GST) tag and the target protein, we designed three primers, attB1-rTEV primer (5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCGAAAACCTGTATTTTCAGGGC-3′), rTEV-merlin primer (5-TCCGAAAACCTGTATTTTCAGGGCATGGCCGGGGCCATCGCTTCCCGC-3′) and attB2-merlin primer (5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTCATCGAGCGAGGCCACGCTGCCGCTCCATCTGCTTTCTATCC-3′).

A PCR product generated by two-step PCR (rTEV-merlin primer and attB2-merlin first and then attB1-rTEV primer and attB2-merlin primer) was cloned into pDEST15, a GST fusion protein vector, according to the manufacturer's instructions and this clone was named pDEST15:merlin341. To improve the efficiency of purification, we modified the vector to include a hexa-His ($His_6$) tag at the *Nde*I site in front of GST sequence using the primers 5′-TATGTCAGGGCACCATCACCATCACCATTCTGGGGCTGC-3′ and 5′-TAGCAGCCCCAGAATGGTGATGGTGATGGTGCCCTGACA-3′. This vector was denoted pHisDEST15:merlin341. Finally, we introduced a stop codon after Ala113 to eliminate the amino acids

extraneous to the FERM domain, using the primers 5′-GAGGAGAAGGAAAGCCTAGTCTTTGGAAGTTCAG-CAG-3′ and 5′-CTGCTGAACTTCCAAAGACTAGCCTT-TCCTTCTCCTC-3′. This resulted in the clone pHisDEST15:merlin313, which was used in all subsequent protein-expression experiments.

### 2.2. Protein purification and crystallization

To overexpress the double-tagged merlin FERM domain, pHisDEST15:merlin313 was introduced into *Escherichia coli* BL21 (DE3) RIL strain (Stratagene). LB medium containing ampicillin (50 mg ml$^{-1}$) was inoculated using 5%(*v*/*v*) of overnight seed culture. After cultivation at 310 K for 3 h, 1 m*M* IPTG was added and cells were cultivated at 295 K for a further 12 h. Cells were harvested by centrifugation at 5000*g* for 20 min, resuspended with 50 m*M* Tris–HCl pH 7.5 (buffer *A*) and disrupted by sonication (Sonifier 450, Branson) for 30 s ml$^{-1}$. The cell lysate was centrifuged at 26 000*g* for 45 min and the soluble supernatant was applied to a glutathione-Sepharose 4B column (Amersham Pharmacia Biotech). After washing the column with 50 m*M* Tris–HCl pH 8.5, 50 m*M* NaCl, the recombinant protein was eluted with buffer *B* (10 m*M* glutathione). The eluent was subjected to a HiPrep 26/10 Desalting column (Amersham Pharmacia Biotech) equilibrated with buffer *A* to remove NaCl and glutathione. The recombinant protein was digested using rTEV protease (Life Technologies) at 283 K in the presence of 0.5 m*M* EDTA and 1 m*M* DTT. After digestion, 300 m*M* NaCl was added to the digested recombinant protein solution and it was passed through a glutathione Sepharose 4B column again to remove uncut full-length fusion protein and the $His_6$-GST tag. To remove rTEV protease and residual tag, 10 m*M* imidazole was added to the flowthrough from the glutathione Sepharose 4B column and this solution was loaded onto an Ni–NTA column (Qiagen) equilibrated with buffer *A* containing 300 m*M* NaCl and 10 m*M* imidazole. The flowthrough of this column was concentrated using a Centriprep YM30 (Amicon), loaded onto a Superdex G75 column (Amersham Pharmacia Biotech) and eluted with buffer *A* containing 300 m*M* NaCl. The fractions containing the merlin FERM domain were collected and concentrated using a Centriprep YM30 for crystallization screening. All the purification steps, except the rTEV digestion, were performed at 277 K. The purified FERM domain contains an additional glycine at the N-terminus arising from the rTEV recognition sequence. After the purification, about 30 mg of pure protein was obtained from 2.8 l of culture.

After screening for crystallization conditions using Crystal Screen and ammonium sulfate Grid Screen (Hampton Research), crystallization conditions were optimized around 0.1 *M* sodium cacodylate pH 6.5 containing ammonium sulfate and dioxane. The sitting-drop vapor-diffusion method was used for all crystallization trials. Drops were formed of 3 µl of protein solution and 3 µl of reservoir buffer and were overlaid with a 1:1 mixture of silicone and mineral oils. Crystallization trays were stored at 294 K. The best crystals were obtained using a 5 mg ml$^{-1}$ protein solution and a buffer containing 56% saturated ammonium sulfate, 2% dioxane and 0.1 *M* sodium cacodylate.

**Table 1**
Data-collection and refinement statistics.

Values in parentheses refer to the highest resolution shell.

| | |
|---|---|
| Experimental data | |
| Space group | $P2_12_12_1$ |
| Unit-cell parameters (Å) | |
| *a* | 87.02 |
| *b* | 89.33 |
| *c* | 96.77 |
| Resolution (Å) | 30–1.80 (1.86–1.80) |
| Mosaicity (°) | 0.69 |
| Unique reflections | 68222 (6875) |
| Redundancy | 3.6 (3.4) |
| Completeness (%) | 95.4 (97.2) |
| $R_{sym}$† | 0.065 (0.622) |
| Average $I/\sigma(I)$ | 16.8 (2.68) |
| Reflections with $I > 3\sigma$ (%) | 70.8 (30.2) |
| Refinement details | |
| Resolution (Å) | 5.0–1.8 (1.847–1.8) |
| Reflections (working) | 66303 (4917) |
| Reflections (test) | 985 (77) |
| $R_{work}$† (%) | 19.3 (26.0) |
| $R_{free}$† (%) | 22.7 (26.8) |
| No. of waters | 862 |
| R.m.s. deviation from ideal geometry | |
| Bonds (Å) | 0.011 |
| Angles (°) | 1.39 |
| Average *B* factor (Å$^2$) | |
| Main chain | 23.5 |
| Side chain | 26.1 |
| Waters | 40.6 |
| Sulfate | 44.3 |

† $R_{sym} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$, $R_{work}$ or $R_{free} = \sum ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \sum_{hkl} |F_{obs}(hkl)|$.

### 2.3. Data collection, structure solution and refinement

The crystal used for data collection was briefly soaked in a solution containing 12.5%(*v*/*v*) glycerol and 56% ammonium sulfate before being transferred to 24% glycerol and 30% ammonium sulfate and frozen by immersion in liquid nitrogen. The data were collected at beamline X9B at NSLS at a wavelength of 0.920 Å under cryoconditions using an ADSC Quantum 4 CCD. The data were indexed and scaled using *HKL*2000 (Otwinowski & Minor, 1997).

The structure was solved by molecular replacement using *AMoRe* (Navaza, 1994). The program *SEAMAN* (Kleywegt, 1996*a*) was used to create a search model based on the radixin structure (PDB code 1gc6), with serines substituted for all non-conserved residues larger than alanine. Manual model rebuilding was performed in *O* (Jones *et al.*, 1991). A combination of *CNS* (Brünger *et al.*, 1998) and *REFMAC* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) was used for refinement, with the final refinement performed using *REFMAC*5 with default values for target stereochemistry (Murshudov *et al.*, 1997). Waters were added using *ARP/wARP* (Perrakis *et al.*, 1999).

## 3. Results and discussion

### 3.1. Crystallization, data collection and structure solution

The FERM domain crystals belong to space group $P2_12_12_1$, with unit-cell parameters $a$ = 87.02, $b$ = 89.33, $c$ = 96.76 Å. After two weeks, the average size of the crystals was 0.2 × 0.2 × 0.1 mm (Fig. 1*a*). The volume of the asymmetric unit (188 040 Å$^3$) suggested the presence of two molecules, with a resultant Matthews coefficient of 2.5 Å$^3$ Da$^{-1}$ and a solvent content of 51.5%. Data with overall completeness of 95.4% were collected from a single frozen crystal and were merged and scaled with an $R_{\rm merge}$ of 0.065 (Table 1).

In order to assess which of the two existing atomic models of FERM domains is closer to merlin, parallel molecular-replacement calculations were performed with a model based on the radixin structure at 2.8 Å without IP3 (PDB code 1gc7; Hamada *et al.*, 2000) and a model based on the moesin structure (PDB code 1efl; Pearson *et al.*, 2000). Although the levels of sequence identity of merlin with radixin or moesin are high (64 and 63%, respectively), a model was constructed from each, with non-conserved residues larger than alanine truncated to serines. These two polyserine models yielded molecular-replacement solutions that were marginally better than those obtained from the complete structures (data not shown). With each model, the molecular-replacement calculations gave two solutions, in agreement with expectations based on crystal density considerations. By most statistical criteria, the moesin-based model provided the best solution for the rotation and translation function, but the radixin-based model provided a better solution after rigid-body refinement. Based on this and the slightly higher sequence similarity with radixin, the radixin-based model was used as the starting point for refinement. During the refinement, non-crystallographic symmetry restraints were not applied, given the relatively high resolution of the data. A combination of *CNS* and *REFMAC* was used to refine the structure, with the final rounds of refinement performed in *REFMAC*5. Maximum-likelihood refinement of the model converged with the statistics reported in Table 1. To determine the extent of model bias, several rounds of refinement were also performed using the moesin-based model. This refinement was discontinued when we were satisfied that the model was not significantly biased by the initial model choice.

### 3.2. Quality of the refined atomic model

The final model consists of two molecules of the FERM domain, 861 water molecules and six sulfate ions. The first 19 amino acids in each of the FERM domain molecules are not visible in the electron density. The refined structure conforms to standard protein stereochemistry, with an r.m.s. deviation from ideal bond lengths of 0.011 Å and only five of the 588 residues of the structure falling into generously allowed regions of the Ramachandran plot (Laskowski *et al.*, 1993). Only a few side chains are not entirely contained within the electron density of a $2F_o - F_c$ $\sigma_A$-weighted map at 1$\sigma$ (Figs. 1*c* and 1*d*). Each monomer contains one *cis*-proline. A limited number of residues exhibit static disorder, but alternate conformations were not refined at this point.

The main-chain temperature factors range from 12.0 to 54.3 Å$^2$, with average values of 22.6 and 24.4 Å$^2$ for chains *A* and *B*, respectively. This similarity is easily rationalized by similar packing of both molecules in the crystal lattice. The temperature factors are generally higher, as expected, in external loops. The exception to this is strand $\beta$5*C*, a terminal strand in a $\beta$-sheet, which is stabilized by hydrogen bonds on one side only. The low *B* values reported in this study reflect the superior quality of the atomic model. This is particularly

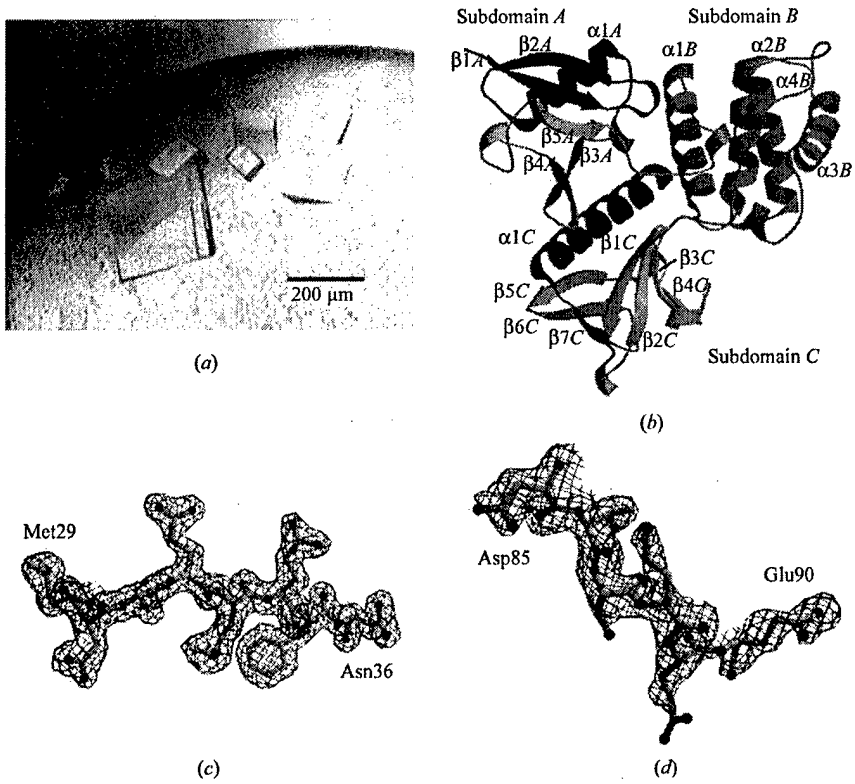



**Figure 1**
(*a*) Typical crystals of the FERM domain of merlin. (*b*) The FERM domain of merlin is shown in a ribbon representation color-ramped from blue to red. The subdomains are labeled as they are described in the text and the main secondary-structural elements are labeled. (*c*) Typical electron density is shown contoured at 1.2$\sigma$ in a $2mF_{\rm obs} - DF_{\rm calc}$ map. (*d*) The electron density for the region with the highest *B* factors is shown contoured at 1.0 $\sigma$ in a $2mF_{\rm obs} - DF_{\rm calc}$ map. Figs. 1(*b*), 2, 3, 4 and 5 were produced using *MOLSCRIPT* (Kraulis, 1991) and *Raster3D*. Figs. 1(*c*) and 1(*d*) were produced with *BOBSCRIPT* (Esnouf, 1997) and *Raster3D* (Merritt & Bacon, 1997).

striking when these values are compared with those found in the radixin models (~64 Å$^2$ for the main chain) or the isolated FERM domain of moesin (PDB code 1e5w; ~70 Å$^2$ for main chain). While this discrepancy may stem from limited resolution in those studies, it is also likely that the refinement protocols may not have been optimal. For example, some of the loops in the moesin FERM domain (PDB code 1e5w) have *B* values in excess of 140 Å$^2$, which corresponds to an unrealistic value of the mean-square displacement $\langle r^2 \rangle$ of nearly 2.0 Å$^2$.

The r.m.s. distance between the C$^\alpha$ atoms of the two molecules, following least-squares overlap, is 0.77 Å. Only a handful of residues, mostly solvent exposed, have different side-chain conformations in the two monomers. The segments with larger differences correlate with areas of higher temperature factors and areas that are involved in crystallographic contacts. The region with the highest discrepancy is the N-terminus of $\alpha 3B$, with the preceding loop and the $3_{10}$-helix. The average r.m.s. coordinate error derived from the program *SIGMAA* in the *CCP*4 suite is 0.11 Å$^2$.

### 3.3. The overall tertiary architecture and comparisons with moesin and radixin

The tertiary structure of the FERM domain of merlin is very close to that of the homologous domains of moesin and radixin (Edwards & Keep, 2001; Hamada *et al.*, 2000; Pearson *et al.*, 2000). The polypeptide chain folds into three clearly identifiable subdomains, each with similarities to known single-domain proteins. These three structural elements were denoted differently for the moesin and radixin structures and we here choose to follow the latter convention, according to which the merlin fragment encompassing residues 20 to approximately 100 is defined as *A*, that including residues 101–215 is denoted *B*, and the third fragment, residues 216–313, is denoted *C* (Fig. 1*b*). As noted by others (Hamada *et al.*, 2000; Pearson *et al.*, 2000), the *A* subdomain has a fold reminiscent of ubiquitin, *B* is similar to the acyl-CoA binding protein and *C* exhibits a fold found in such signaling domains as PTB, PH and EVH1. Merlin is unique in that it has an additional N-terminal extension of 19 amino acids compared with both radixin and moesin. It has been suggested recently based on limited proteolysis experiments that this fragment is disordered (Brault *et al.*, 2001) and our structure fully confirms this prediction. The first amino acid clearly defined in the electron density is Lys20. It is natural to speculate that this portion of merlin is disordered in solution. However, since this region has been shown to be necessary for the proper functioning of merlin and is implicated in actin binding (Brault *et al.*, 2001), it is also possible that it becomes ordered as merlin binds to some effector target.

Least-squares fitting of the merlin FERM domain onto radixin and moesin reveals that the mutual disposition of the three subdomains is relatively well preserved in all three proteins, although concerted shifts of entire subdomains are noticeable albeit small. Such rearrangements affect global comparisons of r.m.s. positional differences, as the latter are

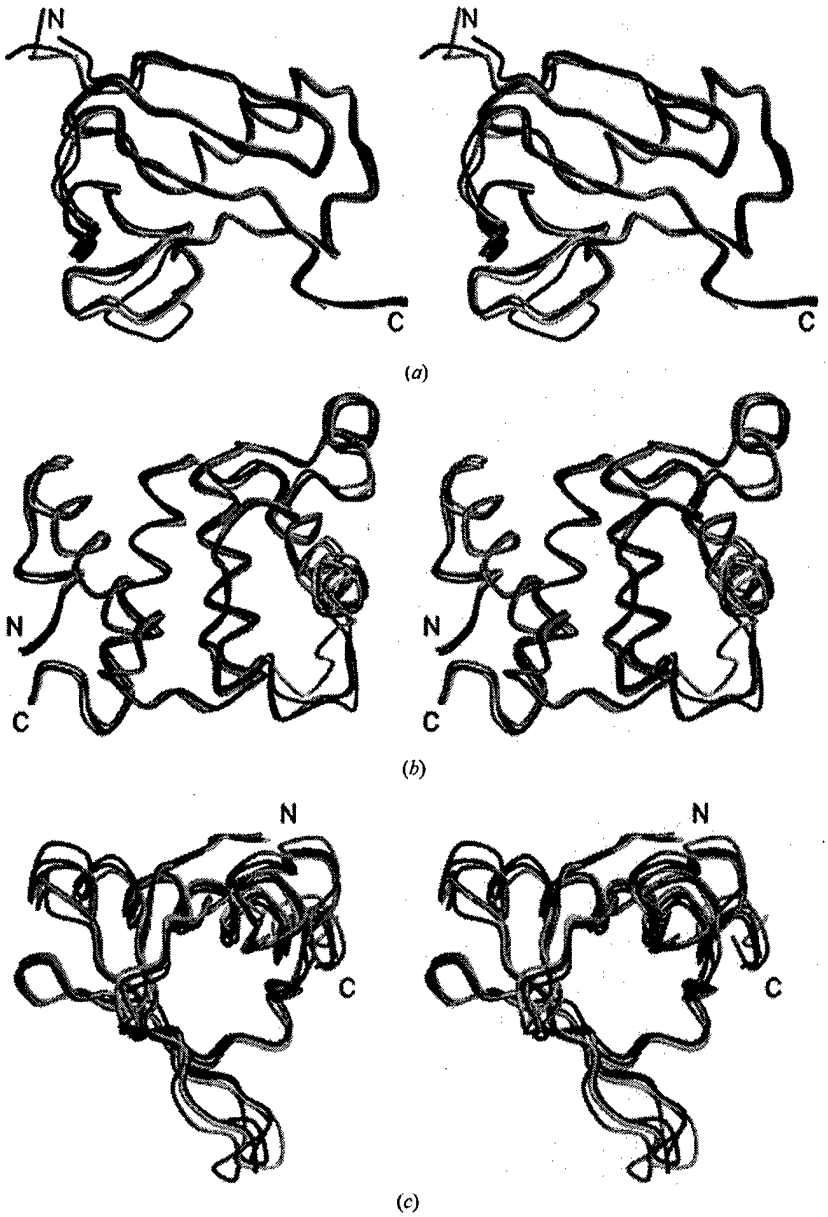



**Figure 2**
Stereoviews of the superpositions of the individual FERM subdomains of merlin, radixin and moesin: subdomain *A* (top), subdomain *B* (center) and subdomain *C* (bottom). In all figures, merlin is blue, radixin is red, 1ef1 moesin is green and 1e5w moesin is gold.

likely to reflect both local changes and global rearrangements. Superpositions of the individual subdomains of the FERM domains of merlin, moesin and radixin are shown in Fig. 2.

To gain a better understanding of the similarities and differences between the FERM domains, the program *LSQMAN* (Kleywegt, 1996*b*) was used to superpose the FERM domains and their individual subdomains. Owing to the high degree of similarity of the two monomers of merlin and for simplicity, values for the *A* monomer are presented here. When the entire ordered portion of the FERM domain of merlin (294 residues) is fitted onto radixin (PDB code 1gc7) or either of the deposited FERM domains of moesin (PDB codes 1efl, which corresponds to the 1.9 Å study, and 1e5w, the 2.7 Å resolution model), the corresponding values of the r.m.s. positional differences between the $C^{\alpha}$ atoms range from 1.8 to 2.0 Å. However, when the subdomains are fitted onto the targets separately and when a few outliers with distances above 3.5 Å are excluded, the values fall dramatically to approximately 0.92 Å for moesin and 0.8 Å for radixin. Subdomain *A*, when fitted onto 1gc7, 1efl and 1e5w, showed r.m.s. differences of 0.65, 0.55 and 0.65 Å, respectively, with 70, 64 and 69 $C^{\alpha}$ atoms within a 1.5 Å distance. The discrepancies occurred consistently around residue 31, the loop comprising amino acids 67–72 and residue 88. For domain *B*, the results were 0.45, 0.65 and 0.51 Å, with 102, 91 and 103 atoms included, respectively. There is a consistent discrepancy, which includes the loop around residue 177. Finally, subdomain *C* exhibits r.m.s. differences of 0.61, 0.67 and 0.70 Å, respectively, with 86, 87 and 77 atoms included and a consistent departure of residues 288–291 from the average. None of the differences are of a magnitude which would suggest a significant biological effect and some can be easily rationalized in terms of crystal contacts.

The high resolution of the present study permits a detailed analysis of the interfaces between the three subdomains, including contributions from the ordered solvent. Two large interfaces contribute to the integrity of the tertiary structure of the FERM domain. The first involves residues from subdomains *A* and *C*. The C-terminal long helix of the *C* subdomain (residues 289–313) packs against two loops of subdomain *A* containing residues 69–76 and 99–103. The face of the helix involved in this interface is largely non-polar and contains Leu306, Leu299 and Leu295; the *A* subdomain contributes Phe100, Trp74 and Val72. Numerous water molecules flank this interface; however, they do not seem to be an integral part of the interface but rather form a typical hydration shell. This specific interface is different in both radixin and moesin because of the single amino-acid deletion which is found in merlin in the loop comprising residues 66–72 and confers a conformational change. As a result, the loop packs significantly closer to the N-terminus of the helix in the *C* subdomain, probably because of a salt bridge formed between Asp70 on one side and Arg291 and Lys289 on the other. Both moesin and radixin lack an aspartate in this position and instead contain bulky aromatics (Phe or Tyr) which push the loop away from the *C* subdomain. The significant difference in the local structure of this loop, as well as the dramatically different amino-acid sequence in this region, suggest that this epitope may be involved in protein–protein interactions unique to merlin.

Another interface is found between subdomains *A* and *B* which, in addition to the intervening loop, interact *via* the first helix of the *B* subdomain, which is wedged between the two subdomains and contributes several hydrophophic side chains such as Ile126, Val122, Phe118 and Phe119. There are also direct hydrogen bonds between the subdomains. This interface is closely packed and lacks any internal water molecules.

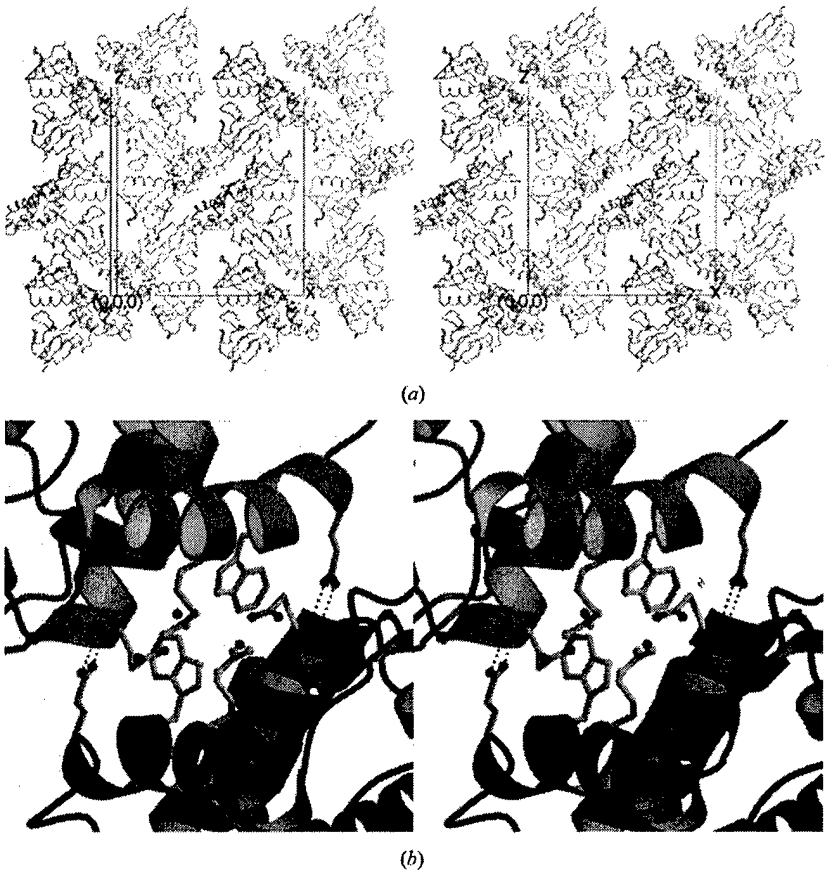


**Figure 3**
Crystal packing of merlin. (*a*) Stereoview of the packing in the unit cell showing the similarities of the packing of the *A* and *C* subdomains for the two merlin monomers. (*b*) Stereoview of the dimer interface. A salt bridge positions the side chains of Glu136 and Arg187 such that they pack against Trp191. Glu194 hydrogen bonds with amide N atoms at the N-terminus of a central helix of subdomain *B*. In both figures monomer *A* is green and monomer *B* is blue.

There is only a small interface between subdomains *C* and *B*. Leu250, located in the loop between $\beta 3C$ and $\beta 4C$, fits into a small pocket formed by Tyr132, Glu215 and Met216. Overall, however, the relative positions of these two subdomains and the entire FERM 'cloverleaf' are defined by the contacts described above and the covalent linkages.

The crystals of the FERM domain of merlin contain two molecules in the asymmetric unit related by a non-crystallographic twofold axis running nearly parallel to the crystallographic *b* axis and between the *B* subdomains of adjacent molecules (Fig. 3*a*). This packing is consistent with a strong maximum in the native Patterson seen at 0.0, 0.5, 0.094 (data not shown), indicating translational non-crystallographic symmetry. The interface between these two molecules, although small (422 Å$^2$), is quite intricate and involves helices $\alpha 2B$ and $\alpha 4B$ in a symmetric arrangement (Fig. 3*b*). The two helices in each molecule interact *via* a salt bridge involving Glu136 and Arg187. Furthermore, the side chains of Arg187 from the two molecules pack tightly against each other and are flanked on each side by the indole rings of the two Trp191 residues. These in turn pack against Glu136 in the neighboring molecule, further stabilizing this contact. At each end of this interface, Glu194 caps otherwise non-bonded backbone amides of residues 136 and 137 at the N-terminus of an $\alpha$-helix, so that each $O^{\epsilon}$ atom accepts a hydrogen bond from one amide N atom. This elegant cap stabilizes the incipient helix immediately downstream of a diprolyl peptide. Finally, two symmetrical pairs of water molecules, each coordinated by at least three hydrogen-bonding partners, add to the stability of this contact. We note that in moesin this general area is involved in the binding of the C-terminal polypeptide and that many of the residues participating in this interface are conserved among the FERM domains, all of which suggests a functional significance.

Other crystal contacts also contribute to the stability of the lattice. Residues 30–36 in subdomain *A* of molecule 1 interact with two loops of subdomain *C* in an adjacent molecule 1, *i.e.* residues 280–282 and 252–255. As both subdomains are roughly at the same *x* and *y* coordinates, this arrangement forms a chain which runs along the *c* axis of the crystal. Interestingly, molecule 2 shows similar contacts and thus the symmetry of the two molecules is broken by the different packing of *B* subdomains against *C* subdomains of molecules in the next layer along the *c* axis. The carboxyl terminus of the *C* subdomain of molecule 1 is buried in the loop which includes residues 169–180 of subdomain *B* of molecule 2, whereas the carboxyl terminus of molecule 2 is wedged between the two *B* subdomains. The structural differences observed between *B* subdomains in moesin, radixin and merlin may reflect, at least in part, the impact of this crystal contact.

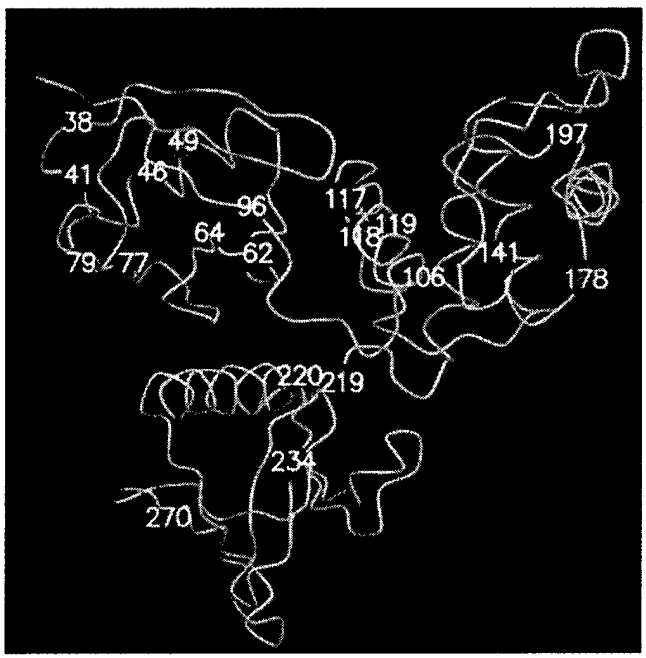



**Figure 4**
Missense mutations of merlin. The distribution of NF2-assocated missense mutations is shown by the presence of a sphere at the $C^{\alpha}$. Red spheres represent a subtitution mutation, purple a deletion and green an insertion. Sites of mutations are labeled.

### 3.4. The NF2-associated missense mutations

The structure of moesin has been used previously to analyze the structural consequences of NF2-associated mutations in merlin. However, given that the current study focuses on the NF2 causal gene product itself, it is proper to address this issue again. Although the most devastating mutations of merlin are nonsense mutations that cause a premature termination of merlin, a number of missense mutations are associated with milder cases of the disease (Gutmann *et al.*, 1998). As can be seen in Fig. 4 and Table 2, 20 of these mutations are distributed throughout the FERM domain, with a slightly higher frequency of mutations in the *A* subdomain. Most of the NF2-assocated mutations occur at sites that are conserved between merlin and other FERM domains, with six of these (Leu46, Phe62, Leu64, Lys79, Phe96 and Ile273) completely conserved among the ERM proteins and protein 4.1. While a number of the NF2-associated mutations are likely to cause critical disruption in the packing of the respective subdomain, the majority of the mutations may impact the subdomain interfaces. This suggests that the specific architecture of the cloverleaf is crucial for the normal function of the protein. None of the mutations occur at the surface interacting with the C-terminal polypeptide of merlin, as predicted from the structure of the moesin complex.

The subdomain interface that is affected by the largest number of NF2-assocated mutations is the *AB* interface. Phe62 is directly involved in the *AB* interface and the mutation of this residue to a serine removes part of the hydrophobic interaction between these two subdomains. The L117I mutation in subdomain *B* is also found at the hydrophobic interface between these two subdomains. The insertion of a leucine in subdomain *A* after residue 49 may also affect the *AB* interface by altering the conformation of the subsequent

**Table 2**
NF2-associated mutations.

HC, hydrophobic core. Sub*A*, subdomain *A*; sub*B*, subdomain *B*; sub*C*, subdomain *C*; ins, insertion; del, deletion.

| Mutation | Structural consequence | Phenotype | Reference |
|---|---|---|---|
| Subdomain *A* mutations | | | |
| E38V | Decreased solubility or impaired interactions | Mild NF2 | Parry *et al.* (1996) |
| W41C | Side-chain packing in sub*A* | Mild NF2 | Welling *et al.* (1996) |
| L46R | HC of sub*A* | Meningiomas | Mérel *et al.* (1995) |
| ins 49L | HC of sub*A* + *AB* hydrogen-bond loss | Mild NF2 | Ruttledge *et al.* (1996) |
| F62S | *AB* hydrophobic interface | Mild or severe NF2 | Scoles *et al.* (1996) |
| L64P | HC of sub*A* | Not reported | Xu & Gutmann (1998) |
| M77V | See text | Intermediate NF2 | Evans *et al.* (2000) |
| K79E | See text | Schwannomas | Sainz *et al.* (1994) |
| del F96 | HC of sub*A* | Severe NF2 | MacCollin *et al.* (1994) |
| Subdomain *B* mutations | | | |
| E106G | *BC* packing (salt-bridge loss in sub*B*) | Severe NF2 | Bourn *et al.* (1994) |
| L117I | *AB* hydrophobic interface | Meningiomas | De Vitis *et al.* (1996) |
| del F118 and/or del F119 | *AB* hydrophobic interface + HC of sub*B* | Severe NF2 | Bourn *et al.* (1995) |
| L141P | Breaks helix and disrupts fold of sub*B* | Not reported | Unpublished† |
| del Q178 | See text | Severe NF2 | Kluwe *et al.* (2000) |
| G197C | Unfavorable conformation in loop | Mild NF2 | Welling *et al.* (1996) |
| Subdomain *C* mutations | | | |
| V219M | *BC* hydrophobic interface | Meningiomas | Mérel *et al.* (1995) |
| N220Y | *AC* hydrophobic interface | Mild NF2 | Ruttledge *et al.* (1996) |
| L234R | HC of sub*C* | Severe NF2 | Jacoby *et al.* (1999) |
| E270G | Loss of salt bridge disrupts sub*B* | Severe NF2 | Kluwe *et al.* (1998) |

† Unpublished mutation found at http://neuro-trials1.mgh.harvard.edu/nf2.

loop. This loop facilitates four inter-subdomain hydrogen bonds, with the extended side chains of Glu58 in subdomain *A* and Gln111 in subdomain *B* hydrogen bonding with the backbone of the adjacent subdomain. Furthermore, deletions of Phe118 and Phe119, together or individually, were also found in NF2 patients. These residues are contained in $\alpha 1B$ and further underscore the functional importance of the *AB* interface.

Several NF2-assocated mutations affect the interfaces between domain *C* and the other two subdomains. Two mutations, V219M and N220Y, are located at the N-terminal strand of subdomain *C*. Their side chains point in opposite directions, with the side chain of 219 pointing towards the *BC* interface and the side chain of 220 in the direction of the *AC* interface. Introduction of a bulkier side chain at either site may disrupt the respective interface. The mutation E106G removes a side chain involved in a salt bridge with Lys209, possibly allowing subdomain *B* to rotate closer to subdomain *C*. In other ERM proteins a serine or alanine is found in this position, making this interaction unique to merlin.

The majority of the other NF2-assocated mutations are most likely to disrupt local packing within their respective subdomains. L46R and L234R introduce a large charged residue in the hydrophobic core of subdomains *A* and *C*, respectively. The L64P substitution and the deletion of Phe96 would create cavities in the hydrophobic core of subdomain *A*, as well as disrupt the local secondary structure surrounding these residues. The G197C substitution occurs at the loop between $\alpha 2B$ and $\alpha 3B$, which requires a backbone conformation that is unfavorable for cysteine ($\varphi = 100$, $\psi = -13°$) and may decrease the solubility of the protein by exposing the side chain to solvent. L141P may destabilize subdomain *B* by inserting a proline into the middle of one of the central helices ($\alpha 2B$). Although this residue is not on the side of the helix that points towards $\alpha 4B$, it is approximately at the position where these two central helices cross each other. The E270G mutation is likely to destabilize the *C* subdomain and alter one of the potential effector-binding sites (see below).

Several of the mutations do not clearly fall into the categories of disrupting subdomain interfaces or subdomain tertiary structure. One such mutation is K79E, which is at the end of $\alpha 4A$. This charge-reversing mutation is very likely to cause the formation of a salt bridge with the neighboring Lys76. Both of these lysines are conserved among ERM proteins. In the merlin structure Lys76 is hydrogen bonded to Tyr66 in $\alpha 3A$, which is also conserved in the ERM family. However, in the radixin structure the homologous lysine extends outward and interacts with the IP3. Although as yet there is no direct evidence that merlin binds inositol phosphates, almost all of the residues responsible for the binding of IP3 in the radixin structure are either conserved in merlin or replaced with functionally equivalent amino acids. The charge reversal caused by the K79E mutation would most likely prevent any inositol phosphates from binding to this pocket.

The potential effect of mutations of residues with solvent-exposed side chains is less clear. The side chain of Met77, mutated to a valine in at least one NF2 case, packs against Phe47 and the mutation may create a destabilizing solvent-accessible depression. Similarly, the substitution of Glu38 for a valine is in a solvent-exposed region. Although this substitution is sterically accommodated, it would place a hydrophobic residue on the surface of the protein, possibly substantially decreasing the solubility of merlin. It is noteworthy that the same type of substitution in hemoglobin causes sickle-cell anemia. The nearby mutation of W41C would affect the local packing of side chains in the area surrounding Glu38. The deletion of Gln178 is discussed below.

### 3.5. Other functional implications

The apparent differences in the biological properties of the various members of the ERM family call for a careful analysis of their respective molecular models. It has been suggested recently that the FERM domain in complex with the C-terminal polypeptide is in a 'dormant' state and that its biological inertness is a product of the occlusion of the relevant epitopes and conformational differences (Edwards & Keep, 2001). This suggestion is based on the 2.7 Å analysis of

the structure of the uncomplexed FERM domain of moesin and on its comparison with the structure of the complexed moesin at 1.9 Å resolution (Pearson *et al.*, 2000). In particular, the loop encompassing residues 260–264 (276–280 in merlin) was found to differ significantly between the two models. We note, however, that the cloverleaf-like fold has some intrinsic flexibility made possible by the interfaces between subdomains. Crystal contacts are sufficient to force minor distortions, but individual domains remain nearly identical within experimental error in their tertiary fold. Although the 276–280 loop in our structure resembles the conformation described by Edwards & Keep (2001), we believe that this does not necessarily constitute proof that the difference is caused by the binding of the C-terminal fragment.

The distribution of residues conserved in moesin, radixin and ezrin but not in merlin can shed light onto the origin of the functional differences between merlin and the ERM proteins. There are 72 such residues and an additional 19 are found in two of the three ERM proteins but not in merlin. Almost all are located at the surface of the protein, although they are not evenly distributed over the surface (Fig. 5). Of these 91 residues unique to merlin, 31 result in a change in the surface electrostatics. Relatively few affect epitopes involved in the binding of the C-terminal polypeptide. The majority of the 91 residues are clustered in three patches, two of which are roughly at a tip of the cloverleaf. One patch is located in each subdomain and therefore the patches will be described as patches *A*, *B* and *C*. These patches are likely to interact with effectors or activators of merlin.

Patch *C* is at the C-terminal end of subdomain *C* and includes resides $\beta5C$–$\beta7C$ and the beginning of $\alpha1C$. All of the merlin-specific residues in this area have their side chains exposed to the solvent and four of them, located on the face of the second $\beta$-sheet in this subdomain, involve a charge change from the ERM consensus sequence. Glu270 and Lys284, mentioned above in the context of the E270G mutations, both constitute a charge change from the other ERM members and are located in patch *C*.

The second patch of residues unique to merlin is found near the tip of subdomain *A* and includes residues found in the distal ends of $\beta1A$ and $\beta2A$, in the following loop and in the N-terminal end of $\alpha1A$. Ezrin has been shown to contain an actin-binding site in this area (Martin *et al.*, 1997). Although the overall net charge of the region is unchanged from the ERM consensus sequence, the local electrostatic footprint is altered by the addition of two acidic and two basic residues, making it unlikely that this serves as an actin-binding site in merlin. Moreover, merlin has been shown to contain an actin-binding site within the first 27 residues, 19 of which are not found in the ERM proteins (Brault *et al.*, 2001). The fact that the E38V and W41C mutations are included in patch *A* makes it more likely that the effects of this protein are manifested by impairing the ability of merlin to bind to effectors or activators.

The subdomain *B* patch contains the beginning of $\alpha4B$ and the loop that precedes it. This region has been called the 'Blue Box' in the *Drosophila* homolog of merlin and has been shown to be vital for the protein's function (LaJeunesse *et al.*, 1998). A comparison of merlin with moesin complexed with its C-terminal fragment reveals that the Blue Box is adjacent to the loop between the *A* and *B* helices of the C-terminal polypeptide. The residues of that fragment that contact the Blue Box are not conserved between merlin and other ERM proteins; thus, the molecular surface that covers the most extended part of the C-terminal polypeptide and the flanking region of merlin is different from the corresponding regions of ezrin, radixin and moesin. This could explain why the activation of merlin is not coincident with any of these proteins. This

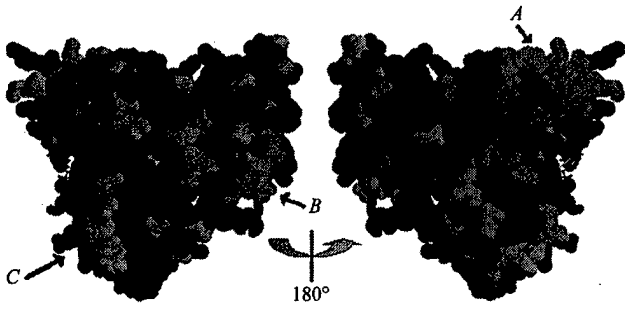



**Figure 5**
The locations of the residues unique to merlin are shown in green on a blue space-filled model. The arrows point to the patches described in the text. In this figure, the C-terminal polypeptide of moesin has been roughly positioned on merlin to indicate where the FERM and C-terminal polypeptide interaction is most likely to occur in merlin. The IP3 of radixin is also included. The image on the left is in the same orientation as Fig. 1(*b*).

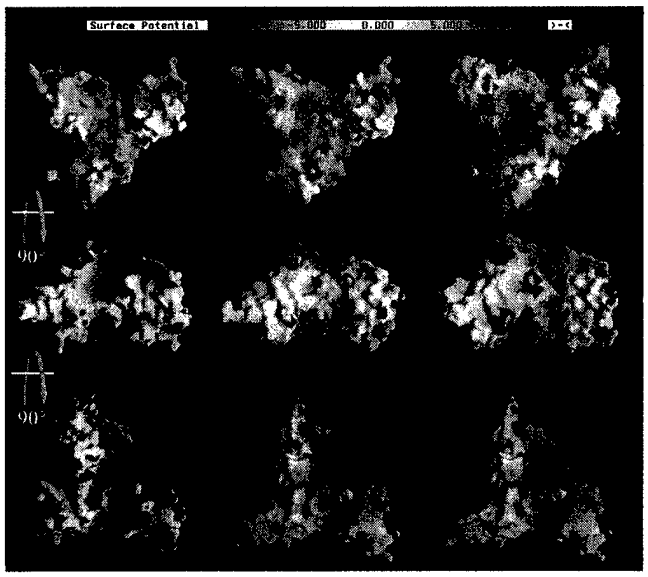



**Figure 6**
Electrostatic potentials generated in *GRASP* (Nicholls *et al.*, 1991) are shown for merlin (left), radixin (middle) and moesin (1ef1) (right). The top views are in the same orientation as in Fig. 1(*b*) and each successive image down the figure has been rotated 90° forward.

is also supported by the NF2 phenotype associated with the deletion of Gln178 located in the Blue Box region. Gln178 is conserved among the ERM members but not in merlin and is adjacent to where the most extended part of the C-terminal polypeptide would be as judged by the moesin complex structure. The nature of this loop would lead one to believe that this loop could rearrange itself to accommodate this deletion without too much difficulty; however, this mutation leads to a severe NF2 phenotype (Kluwe *et al.*, 2000).

Although the patches of residues unique to merlin create local epitopes, the overall molecular surface and electrostatic potential of merlin is similar to that of other members of the ERM family. The largest exception to this is the *AB* interface (Fig. 6). This cleft is much more electronegative than in the other ERM proteins. It is interesting to note that this is the surface that is affected by many of the NF2-associated missense mutations and is roughly flanked by patches *A* and *B*. This leads one to speculate that this region is crucial for the interaction of merlin with effectors or activators.

## 4. Conclusions

We have described the structure of the FERM domain of merlin at 1.8 Å resolution, the highest resolution to date for any of the FERM proteins. As expected, the structure is similar to those of the respective domains in radixin and moesin, but also exhibits interesting differences which may have functional implications. This work sets the stage for more detailed analysis of structure–function relationships in merlin, with the aim of designing ways of either subduing or eliminating the devastating symptoms of neurofibromatosis type 2.

We would like to thank Dr Zbigniew Dauter (NCI, NSLS) for help with data collection and Mary Lewis for technical assistance. The research on merlin in the laboratory is supported by NIH grant HL48807 and CDMRP grant NF000025.

## References

Bourn, D., Carter, S. A., Mason, S., Gareth, D., Evans, R. & Strachan, T. (1994). *Hum. Mol. Genet.* **3**, 813–816.
Bourn, D., Evans, G., Mason, S., Tekes, S., Trueman, L. & Strachan, T. (1995). *Hum. Genet.* **95**, 572–574.
Brault, E., Gautreau, A., Lamarine, M., Callebaut, I., Thomas, G. & Goutebroze, L. (2001). *J. Cell Sci.* **114**, 1901–1912.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
De Vitis, L. R., Tedde, A., Vitelli, F., Ammannati, F., Mennonna, P., Bigozzi, U., Montali, E. & Papi, L. (1996). *Hum. Genet.* **97**, 632–637.
Edwards, S. D. & Keep, N. H. (2001). *Biochemistry*, **40**, 7061–7068.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–143.
Evans, D. G., Huson, S. M., Donnai, D., Neary, W., Blair, V., Newton, V. & Harris, R. (1992). *Q. J. Med.* **84**, 603–618.
Evans, D. G., Sainio, M. & Baser, M. E. (2000). *J. Med. Genet.* **37**, 897–904.
Goutebroze, L., Brault, E., Muchardt, C., Camonis, J. & Thomas, G. (2000). *Mol. Cell Biol.* **20**, 1699–1712.
Gutmann, D. H. (2001). *Hum. Mol. Genet.* **10**, 747–755.
Gutmann, D. H., Geist, R. T., Xu, H., Kim, J. S. & Saporito-Irwin, S. (1998). *Hum. Mol. Genet.* **7**, 335–345.
Gutmann, D. H., Sherman, L., Seftor, L., Haipek, C., Hoang Lu, K. & Hendrix, M. (1999). *Hum. Mol. Genet.* **8**, 267–275.
Hamada, K., Shimizu, T., Matsui, T., Tsukita, S. & Hakoshima, T. (2000). *EMBO J.* **19**, 4449–4462.
Han, B. G., Nunomura, W., Takakuwa, Y., Mohandas, N. & Jap, B. K. (2000). *Nature Struct. Biol.* **7**, 871–875.
Herrlich, P., Morrison, H., Sleeman, J., Orian-Rousseau, V., Konig, H., Weg-Remers, S. & Ponta, H. (2000). *Ann. NY Acad. Sci.* **910**, 106–120.
Jacoby, L. B., MacCollin, M., Parry, D. M., Kluwe, L., Lynch, J., Jones, D. & Gusella, J. F. (1999). *Neurogenetics*, **2**, 101–108.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. (1996*a*). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **32**, 32–36.
Kleywegt, G. J. (1996*b*). *Acta Cryst.* D**52**, 842–857.
Kluwe, L., MacCollin, M., Tatagiba, M., Thomas, S., Hazim, W., Haase, W. & Mautner, V. F. (1998). *Am. J. Med. Genet.* **77**, 228–233.
Kluwe, L., Mautner, V., Parry, D. M., Jacoby, L. B., Baser, M., Gusella, J., Davis, K., Stavrou, D. & MacCollin, M. (2000*a*). *Neurogenetics*, **3**, 17–24.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
LaJeunesse, D. R., McCartney, B. M. & Fehon, R. G. (1998). *J. Cell Biol.* **141**, 1589–1599.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
McClatchey, A. I., Saotome, I., Mercer, K., Crowley, D., Gusella, J. F., Bronson, R. T. & Jacks, T. (1998). *Genes. Dev.* **12**, 1121–1133.
MacCollin, M., Ramesh, V., Jacoby, L. B., Louis, D. N., Rubio, M. P., Pulaski, K., Trofatter, J. A., Eldridge, R., Parry, D., Bove, C., Short, M. P. & Gusella, J. (1994). *Am. J. Hum. Genet.* **55**, 314–320.
Maeda, M., Matsui, T., Imamura, M. & Tsukita, S. (1999). *Oncogene*, **18**, 4788–4797.
Mangeat, P., Roy, C. & Martin, M. (1999). *Trends Cell Biol.* **9**, 187–192.
Martin, M., Roy, C., Montcourrier, P., Sahuquet, A. & Mangeat, P. (1997). *Mol. Biol. Cell*, **8**, 1543–1557.
Martuza, R. L. & Eldridge, R. (1988). *N. Engl. J. Med.* **318**, 684–688.
Meng, J. J., Lowrie, D. J., Sun, H., Dorsey, E., Pelton, P. D., Bashour, A. M., Groden, J., Ratner, N. & Ip, W. (2000). *J. Neurosci. Res.* **62**, 491–502.
Mérel, P., Hoang-Xuan, K., Sanson, M., Moreau-Aubry, A., Bijlsma, E. K., Lazaro, C., Moisan, J. P., Resche, F., Nishisho, I., Estivill, X., Delattre, J. Y., Poisson, M., Theillet, C., Hulsebos, T., Delattre, O. & Thomas, G. (1995). *Genes Chromosomes Cancer*, **13**, 211–216.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murthy, A., Gonzalez-Agosti, C., Cordero, E., Pinney, D., Candia, C., Solomon, F., Gusella, J. & Ramesh, V. (1998). *J. Biol. Chem.* **273**, 1273–1276.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nguyen, R., Reczek, D. & Bretscher, A. (2001). *J. Biol. Chem.* **276**, 7621–7629.
Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Obremski, V. J., Hall, A. M. & Fernandez-Valle, C. (1998). *J. Neurobiol.* **37**, 487–501.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Parry, D. M., MacCollin, M. M., Kaiser-Kupfer, M. I., Pulaski, K., Nicholson, H. S., Bolesta, M., Eldridge, R. & Gusella, J. F. (1996). *Am. J. Hum. Genet.* **59**, 529–539.
Pearson, M. A., Reczek, D., Bretscher, A. & Karplus, P. A. (2000). *Cell*, **101**, 259–270.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Rouleau, G. A., Merel, P., Lutchman, M., Sanson, M., Zucman, J., Marineau, C., Hoang-Xuan, K., Demczuk, S., Desmaze, C., Plougastel, B., Pulst, S. M., Lenoir, G., Bijsma, E., Fashold, R., Dumanski, J., de Jong, P., Parry, D., Eldridge, R., Aurias, A., Delattre, O. & Thomas, G. (1993). *Nature (London)*, **363**, 515–521.

Ruttledge, M. H., Andermann, A. A., Phelan, C. M., Claudio, J. O., Han, F. Y., Chretien, N., Rangaratnam, S., MacCollin, M., Short, P., Parry, D., Michels, V., Riccardi, V. M., Weksberg, R., Kitamura, K., Bradburn, J. M., Hall, B. D., Propping, P. & Rouleau, G. A. (1996). *Am. J. Hum. Genet.* **59**, 331–342.

Sainz, J., Huynh, D. P., Figueroa, K., Ragge, N. K., Baser, M. E. & Pulst, S. M. (1994). *Hum. Mol. Genet.* **3**, 885–891.

Scoles, D. R., Baser, M. E. & Pulst, S. M. (1996). *Neurology*, **47**, 544–546.

Scoles, D. R., Huynh, D. P., Chen, M. S., Burke, S. P., Gutmann, D. H. & Pulst, S. M. (2000). *Hum. Mol. Genet.* **9**, 1567–1574.

Sherman, L., Xu, H. M., Geist, R. T., Saporito-Irwin, S., Howells, N., Ponta, H., Herrlich, P. & Gutmann, D. H. (1997). *Oncogene*, **15**, 2505–2509.

Trofatter, J. A., MacCollin, M. M., Rutter, J. L., Murrell, J. R., Duyao, M. P., Parry, D. M., Eldridge, R., Kley, N., Menon, A. G., Pulaski, K., Haase, V., Ambrose, C., Munroe, D., Bove, C., Haines, J., Martuza, R., McDonald, M., Seizinger, N., Short, M. P., Buckler, A. & Guzella, J. (1993). *Cell*, **72**, 791–800.

Tsukita, S., Oishi, K., Sato, N., Sagara, J. & Kawai, A. (1994). *J. Cell. Biol.* **126**, 391–401.

Tsukita, S. & Yonemura, S. (1997). *Curr. Opin. Cell. Biol.* **9**, 70–75.

Tsukita, S., Yonemura, S. & Tsukita, S. (1997). *Trends Biochem. Sci.* **22**, 53–58.

Turunen, O., Sainio, M., Jaaskelainen, J., Carpen, O. & Vaheri, A. (1998). *Biochim. Biophys. Acta*, **1387**, 1–16.

Welling, D. B., Guida, M., Goll, F., Pearl, D. K., Glasscock, M. E., Pappas, D. G., Linthicum, F. H., Rogers, D. & Prior, T. W. (1996). *Hum. Genet.* **98**, 189–193.

Xu, H. M. & Gutmann, D. H. (1998). *J. Neurosci. Res.* **51**, 403–415.

Structure, Vol. 11, 459–468, April, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0969-2126(03)00052-2

# PDZ Tandem of Human Syntenin: Crystal Structure and Functional Properties

**Beom Sik Kang,[1,4] David R. Cooper,[1,4] Filip Jelen,[2,4] Yancho Devedjiev,[1] Urszula Derewenda,[1] Zbigniew Dauter,[3] Jacek Otlewski,[2] and Zygmunt S. Derewenda[1,*]**

[1]Department of Molecular Physiology and Biological Physics and
The Cancer Center
University of Virginia
Charlottesville, Virginia 22908

[2]Laboratory of Protein Engineering
Institute of Biochemistry and Molecular Biology
University of Wroclaw
Tamka 2
50-137 Wroclaw
Poland

[3]Synchrotron Radiation Research Section
Macromolecular Crystallography Laboratory
NCI
Brookhaven National Laboratory
Upton, New York 11973

## Summary

**Syntenin, a 33 kDa protein, interacts with several cell membrane receptors and with merlin, the product of the causal gene for neurofibromatosis type II. We report a crystal structure of the functional fragment of human syntenin containing two canonical PDZ domains, as well as binding studies for full-length syntenin, the PDZ tandem, and isolated PDZ domains. We show that the functional properties of syntenin are a result of independent interactions with target peptides, and that each domain is able to bind peptides belonging to two different classes: PDZ1 binds peptides from classes I and III, while PDZ2 interacts with classes I and II. The independent binding of merlin by PDZ1 and syndecan-4 by PDZ2 provides direct evidence for the coupling of syndecan-mediated signaling to actin regulation by merlin.**

## Introduction

Syntenin is a 298 residue long cytosolic protein, originally identified as a molecule linking syndecan-mediated signaling to the cytoskeleton [1]. Subsequently, syntenin was also found to play a role in protein trafficking [2, 3], cell adhesion [4], and activation of the transcription factor Sox4 [5]. Of particular medical interest is the recent report that syntenin is overexpressed in breast and gastric cancer cells and promotes their migration and metastasis [6]. The diverse biological functions of this protein are a result of its interactions with numerous targets. There are currently at least ten putative binding partners reported for syntenin, including IL-5 receptor $\alpha$ subunit (IL5R$\alpha$) [5], neuroglian [7], proTGF-$\alpha$ [3], glutamate receptors [8], neurofascin [7], syndecan-4 [1], ephrin B [9, 10], ephrin A7 [9], PTP-$\eta$ [11], neurexin I [12], and merlin [13]. All the binding partners of syntenin are receptors except for merlin, a cytosolic tumor repressor that is a product of the causal gene for type II neurofibromatosis (NF) [14]. Merlin belongs to the protein 4.1 superfamily, which also includes ezrin, moesin, and radixin, and like its homologs, it binds actin [15].

Based on amino acid sequence analyses, syntenin was predicted to contain a tandem of PDZ domains (PDZ1 and PDZ2) preceded by an N-terminal fragment of 112 amino acids of an unknown structure. PDZ domains are ubiquitous signaling domains, with over 400 distinct copies in the human genome [16, 17], which mediate protein-protein interactions. They may occur in proteins harboring other domains, such as SH2, RGSL, PH, DH, or GUK, but are also found in proteins that contain no other domains: an extreme example, MUPP, is a scaffolding protein with 13 PDZ domains [18]. Through the PDZ domains, signaling proteins bind to receptors, channels, and other targets, often functioning as membrane-associated scaffolds for the assembly of signaling complexes. Finally, PDZ-containing proteins provide a means for subcellular targeting of their partners, as exemplified by the function of Lin-2/CASK [19], Lin-10/MINT1 [20], and GRIP [21, 22].

PDZ domains are structurally conserved modules, about 90 amino acids in length, with a distinct fold of six $\beta$ strands and two $\alpha$ helices [23, 24]. In most cases, they recognize C-terminal sequence motifs of target proteins and bind these peptides in a pocket between the $\beta$2 strand and $\alpha$2 helix. The PDZ domains are typically grouped into three classes depending on the target tripeptides: class I (-S/T-X-$\phi$), class II (-$\phi$-X-$\phi$), and class III (-D/E-X-$\phi$) [17]. Examples outside this paradigm are well documented, and some PDZ domains show degenerate specificity [25]. It has also been reported that interaction between adjacent PDZ domains may modulate peptide binding, further complicating the picture. For example, the PDZ1-PDZ2 tandem within PSD-95 appears to have different binding properties compared to its isolated PDZ domains [16].

The multitude of syntenin's putative partners, which belong to all three classes of target proteins, suggests that its PDZ domains may also exhibit degenerate specificities. Furthermore, it has been reported that the two domains function in a cooperative fashion: for example, isolated PDZ1 and PDZ2 apparently fail to bind merlin- and IL5R$\alpha$-derived peptides, whereas binding was reported for full-length protein [5, 13]. A requirement for the complete tandem was also reported for interaction with PTP-$\eta$ [11] and proTGF-$\alpha$ [3], whereas syndecan-2 was reported to bind to PDZ1-PDZ2 or PDZ2-PDZ2 tandem, but neither the isolated domains nor PDZ1-PDZ1 [12].

In order to explain the molecular basis for the ob-

*Correspondence: zsd4n@virginia.edu
[4]These authors contributed equally to this work.

Key words: PDZ; syndecan; merlin; schwannoma; cancer; crystallography; calorimetry

served properties of syntenin, we have initiated a systematic study aimed at characterizing syntenin's molecular structure and mechanism of action. Here, we report the crystal structure of the intact PDZ tandem, residues 113–273, refined at 1.94 Å resolution. The structure reveals two PDZ domains that have fully solvent-accessible peptide binding pockets. We also report the results of rigorous biophysical binding assays for isolated PDZ domains, the PDZ tandem, and full-length syntenin, with peptides derived from three selected putative binding partners for syntenin: IL5Rα, syndecan-4, and merlin. These data reveal that the binding properties of syntenin are a result of the independent binding events of the two PDZ domains, whose specificities show clear degeneracy. The merlin-derived octapeptide shows the highest affinity for syntenin, and a distinct selectivity for PDZ1. This result reaffirms that merlin is a physiologically important partner for syntenin.

## Results and Discussion

### Model Quality and Structure Overview

The structure was solved with a three-wavelength MAD experiment, using a SeMet-labeled protein crystal. The model, refined at 1.94 Å resolution to a crystallographic R value of 18.4% ($R_{free}$ 22.7%), contains a noncrystallographic dimer of tandems in the asymmetric unit and a total of 325 residues (Table 1; Figure 1). The only syntenin residue not included in the model is the C-terminal phenylalanine of the second monomer. The refined structure conforms well to standard protein stereochemistry, with rms deviation from ideal bond lengths of 0.015 Å, and with only 2 residues falling into disallowed regions of the Ramachandran plot as judged by MolProbity [26]. Only seven side chains are not visible in the $\sigma_A$-weighted $2mF_{obs} - DF_{calc}$ electron density map contoured at 1σ (Figure 1C), and their occupancies were set to zero. The average isotropic temperature factor (B) for main chain atoms is 20.6 Å$^2$, with the highest temperature factors (~50 Å$^2$) associated with the linker peptides and the C-terminal end of the α2 helix of PDZ1, all of which are nonetheless clearly visible.

The crystallized fragment of syntenin contains two PDZ domains conjoined by a short linker (Arg193-Pro194-Phe195-Glu196; Figure 1C). Like other domains from this superfamily, the syntenin PDZ modules show a typical fold with two opposing antiparallel β sheets capped by two α helices. Each domain has at least one β strand that is partly contained in both sheets. In the crystal, the two PDZ tandems of syntenin are arranged in a head-to-tail fashion, related by a noncrystallographic dyad, giving the contents of the asymmetric unit the appearance of a four-leaf clover. Interestingly, the linker residue Arg193 forms a salt bridge with Glu240 in PDZ2 and forms a hydrogen bond via its $N^{\delta}$ with a main chain carboxyl group of Phe154 in PDZ1. This may help to explain why Arg193 falls into a disallowed region of the Ramachandran plot. Superposition of the monomers reveals that there is a slight difference in the angle between the two PDZ domains in the two monomers, explaining why the dimer is noncrystallographic. This suggests that the linker has considerable intrinsic flexibility in solution.

We note that the interaction between the two PDZ domains within a monomer is less extensive than the intermolecular PDZ1-PDZ2 interface. A total of 446 Å$^2$ of solvent-accessible surface is buried by each intermolecular interaction. Furthermore, this interface is fairly intimate, with a number of hydrogen bonds between the two domains. The few solvent molecules that are at this interface mediate contacts between the two domains. In contrast, there are no direct hydrogen bonds between the PDZ domains within a monomer. Both putative peptide binding grooves of syntenin are located on the same face of the tandem monomer and are completely exposed to the solvent, suggesting that syntenin has two distinct and functional peptide binding sites.

### Structure of the PDZ Domains of Syntenin

A structural comparison of the two PDZ domains of syntenin reveals that, in spite of a modest level of sequence identity (26%), they are structurally very similar to each other, with an rms deviation of 1.2 Å on Cα atoms. In both domains, the fragment equivalent to the signature GLGF loop involved in the terminal carboxylate binding deviates from the paradigm by an insertion of a basic residue after the initial glycine (Arg in PDZ1 and His in PDZ2). Such insertions are rarely found in PDZ domains, but they do not seem to disturb the cluster of main chain amides that coordinate the incoming carboxylate of the target peptide. Typically, a Lys or Arg located 4 or 5 residues prior to this loop assists in peptide binding via a water-mediated hydrogen bond. Both of syntenin's PDZ domains have a lysine 4 residues before the initial glycine.

In spite of these similarities, there are some notable differences between the two domains. The most apparent is the length of the β2-β3 loop. When compared to PDZ2, where this loop is shorter than in most other PDZ domains, PDZ1 contains an insertion of 4 residues in the β2-β3 loop (KSIDNGIF versus KN---GK).

Furthermore, in PDZ1, the peptide binding groove is narrower as compared to PDZ2 or other PDZ domains (Figure 2A). This is best illustrated by comparing the distance from the beginning of α2 to the β2 strand of PDZ2, to the corresponding distance in PDZ1, which is 1.8 Å shorter.

The electrostatic potential surrounding the peptide binding groove is another significant difference between the two PDZ domains. The peptide binding surface of PDZ1 is predominately positively charged, surrounded by 3 residues (Lys124, Arg128, and Lys130) from β2 and 2 residues (His175 and Lys179) from α2. Other basic residues flank this region. PDZ2 lacks any clusters of positively charged residues, with His208 as the only charged side chain that extends over the peptide binding groove (Figure 2B).

As a peptide binds to PDZ domains, it mimics an additional antiparallel strand in the sheet containing β2. The position of the β2 strand in both of syntenin's PDZ domains is consistent with this mechanism, with the amino and carboxyl groups of Leu129 and Phe211 available for hydrogen bonding. This type of arrangement dictates that the terminal side chain of the peptide faces the interior of the binding pocket. Both PDZ domains

Table 1. Data Collection and Refinement Statistics

| | Edge | Peak | Remote |
|---|---|---|---|
| **Data Collection Statistics** | | | |
| Wavelength | 0.97946 | 0.97900 | 0.97133 |
| Resolution (Å) | 30.0–1.94 (2.01–1.94)[a] | 30.0–1.94 (2.01–1.94)[a] | 30.0–1.94 (2.01–1.94)[a] |
| Total reflections | 84,171 | 85,123 | 86,109 |
| Unique reflections | 22,557 | 22,608 | 23,095 |
| Redundancy | 3.7 | 3.8 | 3.7 |
| Completeness (%) | 97.4 (75.1) | 97.1 (71.9) | 98.8 (90.2) |
| $R_{sym}$ (%)[b] | 4.9 (25.9) | 6.2 (31.2) | 4.8 (31.1) |
| Average I/σ (I) | 20.2 | 23.1 | 19.7 |
| **Phasing Statistics** | | | |
| Phasing power,[c] iso/ano | 0.37/0.27 | 0.23/0.35 | –/0.31 |
| $R_{Cullis}$,[d] iso/ano | 0.43/0.69 | 0.71/0.57 | –/0.66 |
| Overall figure of merit (acentric): 0.68 | | | |
| **Refinement Statistics** | | | |
| Resolution (Å) | 30.0–1.94 (1.99–1.94) | | |
| Reflections (working) | 21,926 (1,560) | | |
| Reflections (test) | 1,182 (79) | | |
| $R_{work}$ (%)[e] | 18.4 (24.0) | | |
| $R_{free}$ (%)[e] | 22.7 (25.2) | | |
| Number of waters | 254 | | |
| Rms deviation from ideal geometry | | | |
| Bonds (Å) | 0.015 | | |
| Angles (°) | 1.80 | | |
| Average B factor (Å$^2$) | | | |
| Main chain | 20.6 | | |
| Side chain | 26.3 | | |
| Waters | 47.0 | | |

[a]The numbers in parentheses describe the relevant value for the highest resolution shell.
[b]$R_{sym} = \Sigma||I_i - \langle I\rangle|/\Sigma I$, where $I_i$ is the intensity of the i-th observation and $\langle I\rangle$ is the mean intensity of the reflections. The values are for unmerged Friedel pairs.
[c]Phasing power = <[|Fh(calc)|/phase-integrated lack of closure] >
[d]$R_{Cullis}$ = <phase-integrated lack of closure>/<|Fph − Fp|>
[e]$R = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, crystallographic R factor, and $R_{free} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, where all reflections belong to a test set of randomly selected data.

have a substantial hydrophobic pocket near the binding loop that could accommodate any of the large hydrophobic side chains. The walls and floor of the peptide binding groove in both PDZ domains are lined with hydrophobic residues. Neither PDZ domains has a histidine at the first position of α2, as is found in typical class I PDZ domains. In syntenin, this position is occupied in PDZ1 and PDZ2 by Ser171 and Asp251, respectively. The side chains of both of these residues hydrogen bond to main chain amides at the end of β2. Overall, both of the PDZ domains of syntenin appear to be suitable for peptide binding, although the structural differences suggest diverse specificities.

### Stability Studies

Although the crystal structure indicates that both PDZ domains of syntenin are capable of binding peptides, many previously reported binding studies using isolated PDZ1 and PDZ2 domains have failed. In order to better assess the feasibility of performing binding assays with isolated PDZ1 or PDZ2, we conducted stability studies using chemical denaturation monitored by circular dichroism (Figure 3). Full-length syntenin, the PDZ tandem, and isolated PDZ1 (residues 113–193) and PDZ2 (residues 197–273) were used in the assay. Surprisingly, we found that the isolated PDZ1 and PDZ2 have significantly different stabilities: the free energy of unfolding, $\Delta G_{un}$, for PDZ1 is −3.2 kcal/mol, whereas for PDZ2 it is −4.8 kcal/mol, putting it closer to the average of 5–15 kcal/mol observed for globular proteins. Based on these values, the expected denaturation of the tandem, as simulated by combining single domain transitions, should be less cooperative, with a $\Delta G_{un}$ of −2.07 kcal/mol (Figure 3, insert). However, the experimental unfolding of the tandem follows a cooperative, two-state profile, with a $\Delta G_{un}$ of −4.1 kcal/mol, suggesting that the domains are associated into a single entity undergoing cooperative denaturation. As already noted, PDZ1 of one molecule interacts with PDZ2 of the adjacent one in the crystal structure. The large buried interface suggests that the interaction could be physiologically relevant, as indeed self-association has been reported for syntenin [7]. It is also possible that the two domains interact in this way within a monomer, and that the crystal structure corresponds to a domain-swapped conformation. Whereas our data are consistent with weak, albeit identifiable domain-domain interactions in syntenin, further work will be required to probe this issue.

Finally, we note that the full-length protein also unfolds in a highly cooperative manner, and shows signifi-


Figure 1. The Structure of the PDZ Tandem of Syntenin

(A) A stereo Cα trace with every tenth α carbon represented as a sphere and every twentieth α carbon labeled, and colored from blue to red as a function of residue number.

(B) Ribbon diagram of the asymmetric unit colored by B factors. B factors are represented with low values (12 Å$^2$) colored blue and high values (43 Å$^2$) colored red.

(C) Experimentally determined electron density map of the linker region contoured at 1σ. Residues 189–201 are shown for each monomer. Figures were made using MOLSCRIPT [47] (A and B) and POVSCRIPT+ (http://people.brandeis.edu/~fenn/povscript/) (C) and rendered with RASTER3D.

cantly higher stability ($\Delta G_{un}$ of −6.4 kcal/mol) than any of the other constructs. This result may imply that the N-terminal fragment of syntenin is folded and plays a structural role in the protein, possibly interacting with the PDZ domain(s), so that full-length syntenin denatures as a single entity.

### Binding Properties of Syntenin and of Isolated PDZ Domains

In the case of full-length syntenin and the PDZ tandem, we were interested in both the affinities and the stoichiometries of peptide binding. With that in mind, we carried out binding assays using isothermal titration calorimetry

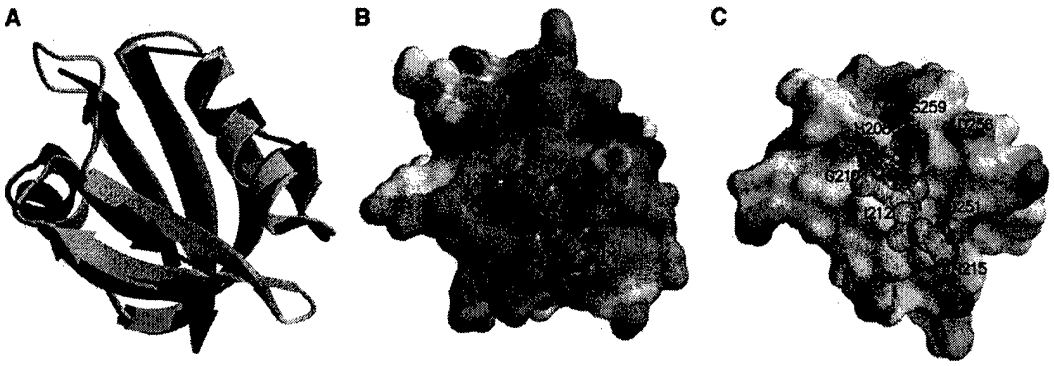



Figure 2. A Comparison of PDZ1 and PDZ2 Domains of Syntenin

(A) Superposition of the two PDZ domains of syntenin. PDZ1 is gold and PDZ2 is blue. The α2 helices have been superposed to show the similarity of the fold, yet emphasize the differences of the peptide binding groove. The same orientation is used for all three figures.

(B) The peptide binding surface of PDZ1. The electrostatic potential surface is shown with select residues that surround the peptide binding groove labeled. A superposed C-terminal CRIPT-derived peptide from the structure of PSD-95 (1BE9 [23]) is shown semitransparent, with side chains represented as cyan spheres in the β carbon position. The approximate locations of the $P_0$, $P_{-1}$, and $P_{-2}$ binding pockets are indicated by gold, pink, and green circles, respectively.

(C) The peptide binding surface of PDZ2 represented as described in (B). Figures were made using MOLSCRIPT [47] (A) and POVSCRIPT+ (http://people.brandeis.edu/~fenn/povscript/) (B and C) and rendered with RASTER3D [48]. Electrostatic potentials were calculated in GRASP [49].

(ITC). The assays were conducted with the following hexapeptides derived from three putative targets of syntenin, each belonging to one of the three canonical classes: LEDSVF (IL5Rα) representing class I, TNEFYA (syndecan-4) from class II, and AFFEEL (merlin) from class III. We found that all peptides bind to full-length syntenin and to PDZ tandem with dissociation constants ($K_d$) in the low μM range (Figure 4; Table 2). Interestingly, the IL5Rα-derived peptide shows a stoichiometry of 2:1 for the tandem, but only 1:1 for full-length syntenin, whereas all other measurements indicate equimolar complexes. This is further evidence that suggests a functional role for the N-terminal domain.

To assess whether residues upstream of the C-terminal hexapeptide contribute to binding, we performed the assays with octapeptides for merlin and IL5Rα. The results are very similar for IL5Rα, but the merlin octapeptide binds an order of magnitude more tightly than the corresponding hexapeptide, indicating the functional importance of residues -6 and -7.

The determination of $K_d$ values for isolated domains by ITC proved difficult, because PDZ1 aggregated at the required high concentration and isolated PDZ2 (residues 197–273) was prone to oligomerization (data not shown). To overcome the aggregation problems, we used a fluorescence-based approach using dansylated hexapeptides, allowing for significantly lower protein concentration [27]. We also changed the PDZ2 construct to residues 197–298, which includes syntenin's natural C terminus. To assess whether either the technique or

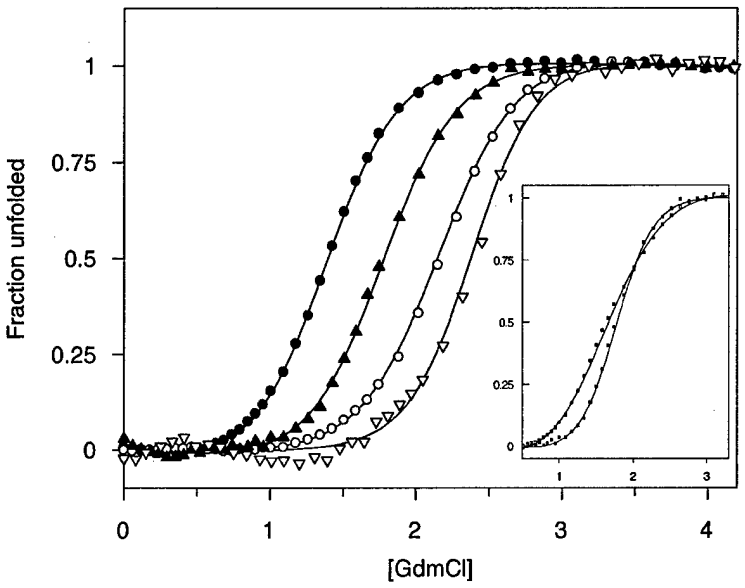



Figure 3. Stability of Syntenin Constructs

GdmCl-induced unfolding of PDZ1 (●), PDZ2 (○), tandem of PDZ domains (▲), and full-length (▽) of syntenin. Measurements were performed in 25 mM Tris, 50 mM NaCl (pH 7.4). Transitions were monitored by the changes of the CD signal at 222 nm. Data were normalized as "fraction unfolded" and fitted to the equation in the text. Insert: combined single domain transitions (□) and tandem of PDZ domains transition (■).

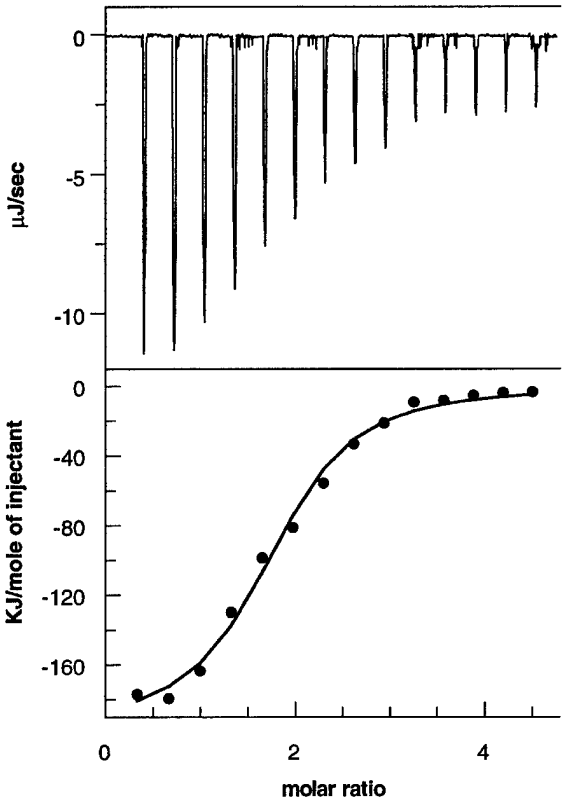



Figure 4. Representative Calorimetric Titration of PDZ Tandem of Syntenin with LEDSVF Peptide

Top: raw heat data corrected for base drift, obtained from 14 consecutive injections of 11.2 mM LEDSVF peptide into a sample cell (1,250 μl) containing 140 μM PDZ tandem of syntenin.

Bottom: the binding isotherm created by plotting the areas under the peaks against the molar ratio of the peptide added to the PDZ tandem present in the cell and the fit line to the model of independent sites. The heats of mixing (dilution) have been subtracted.

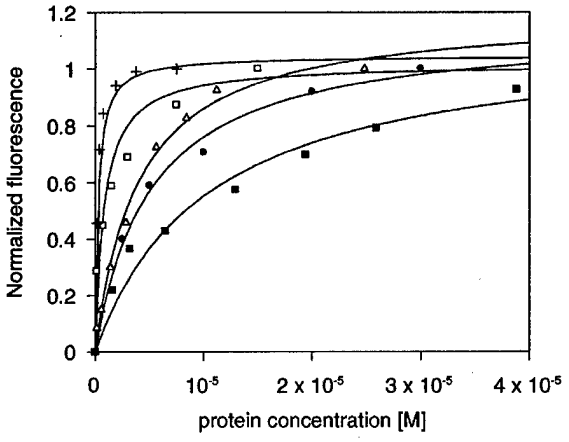



Figure 5. Binding of Dansyl-Labeled Peptides to Syntenin

Binding of dansyl-RVAFFEEL to PDZ1 (+), dansyl-AFFEEL to PDZ1 (●), dansyl-LEDSVF to PDZ1 (■), and PDZ2 (□) and dansyl-TNEFYA to PDZ2 (Δ). Data were normalized as "fraction bound," so that the initial fluorescence was zero and the fluorescence at saturation was equal to unity.

peptide dansylation influenced the observed affinities, we included the PDZ tandem in our measurements as a control. We found that the fluorescence data agree well with the calorimetric titrations (Figure 5; Table 3). The IL5Rα peptide binds to both PDZ domains with similar $K_d$ values in the mid-μM range, in agreement with the 2:1 stoichiometry observed by ITC. The merlin-derived peptide shows no significant affinity toward PDZ2, but binds to PDZ1, to the tandem and the full-length protein, with almost identical $K_d$ values in the sub-μM range. The merlin octapeptide binds to PDZ1 with significantly higher affinity than the hexapeptide, in agreement with the ITC results. The syndecan-4 peptide interacts exclusively with PDZ2, with an affinity virtually identical to those observed for the PDZ tandem and full-length syntenin. This result is again consistent with the 1:1 stoichiometry determined by ITC. However, it is in conflict with the previously reported 2:1 stoichiometry for the whole C-terminal domain of syndecan-2 [12].

It has been suggested previously that the N-terminal fragment of syntenin plays a regulatory function. For example, the association of PTP-η with syntenin was shown to be regulated by tyrosine phosphorylation within this fragment, with phosphorylation preventing the association [11]. This indicates that the N-terminal fragment may, at least under some conditions, regulate the availability of at least one of syntenin's peptide binding pockets. Our data do not provide a clear answer in

Table 2. Isothermal Calorimetry of Syntenin Interactions

| | LEDSVF (IL5Rα) | ETLEDSVF (IL5Rα) | AFFEEL (Merlin) | RVAFFEEL (Merlin) | TNEFYA (Syndecan) |
|---|---|---|---|---|---|
| PDZ Tandem | | | | | |
| $K_d$ | 43.8 μM | 32.2 μM | 11.6 μM | 200 nM | 2.9 μM |
| n | 2.26 | 2.07 | 0.94 | 1.12 | 1.11 |
| ΔH | −2.1 kJ | −7.6 kJ | −7.1 kJ | −3.1 kJ | −5.4 kJ |
| Full-Length | | | | | |
| $K_d$ | 19.5 μM | 10.1 μM | 8.9 μM | 869 nM | 2.5 μM |
| n | 1.09 | 1.12 | 1.3 | 1.16 | 1.14 |
| ΔH | −4.7 kJ | −8.87 | −8.9 kJ | −7.3 kJ | −9.8 kJ |

Dissociation constants, stoichiometries, and enthalpies for the interactions of the IL5Rα-, merlin-, and syndecan-derived peptides, with the PDZ tandem of syntenin and full-length protein, determined by ITC. Representative data are shown for experiments that were conducted at least twice for each interaction.

Table 3. Dissociation Constants of Syntenin Interactions by Fluorimetric Titrations

| | DNS-EDSVF (IL5Rα) | DNS-FFEEL (Merlin) | DNS-VAFFEEL (Merlin) | DNS-NEFYA (Syndecan) |
|---|---|---|---|---|
| PDZ-PDZ | 18.5 μM (±3.2) | 6.8 μM (±1.1) | 495 nM (±55) | 2.1 μM (±0.3) |
| PDZ1 | 10.6 μM (±0.8) | 5.0 μM (±0.7) | 268 nM (±28) | >1 mM |
| PDZ2 | 1.9 μM (±0.3) | >0.6 mM | >1 mM | 2.3 μM (±0.5) |

Dissociation constants for the interactions of the IL5Rα-, merlin-, and syndecan-derived dansylated peptides, with the PDZ tandem of syntenin and isolated PDZ domains, determined by fluorimetric titration. Estimates of error are derived from experimental data.

this regard. The ITC data show that the IL5Rα peptide binds to only one PDZ domain of full-length syntenin, but to both in the tandem. Merlin, which binds only to PDZ1, does so within the context of full-length syntenin as well, suggesting that it is the PDZ2 that is occluded by the N-terminal domain. However, the syndecan-4 peptide is selective for PDZ2 and also binds to full-length syntenin. Further experiments will be necessary to resolve this inconsistency.

Taken together, the structural and binding data indicate that the two domains within the syntenin PDZ tandem function independently. Each domain shows degenerate specificity, so that PDZ1 binds peptides from merlin and IL5Rα, whereas PDZ2 shows affinity toward IL5Rα and syndecan-4. Although our data are internally consistent and reproducible, they are in conflict with some reports in the literature that claim individual PDZ domains are incapable of binding peptides.

### Similarities to Other PDZ Domains

As the number of known PDZ domains grows and their importance in a myriad of cellular events becomes evident, numerous attempts have been made to elucidate the factors that govern their specificity. High-resolution crystal structures and solution NMR structures have now been determined for a number of PDZ domains that were classified into distinct groups. It is clear that the overall fold of the domain is well conserved, and the specificity is governed by subtle structural and amino acid sequence variation. The application of generalized rules for governing PDZ domain specificity is complicated by those PDZ domains that show degenerate specificity for more than one archetypal class of peptide. Syntenin is one of the examples of this growing group. Both of syntenin's PDZ domains fit the overall fold of PDZ domains well, with an average rms deviation from the known X-ray structures of 1.4 Å and 1.1 Å for PDZ1 and PDZ2, respectively.

Syntenin, with its tandem PDZ structure, appears to have been well conserved during evolution. The rat and mouse syntenins are virtually identical to the human protein. Recently, the jumbo tiger shrimp *Penaeus monodon* was reported to contain a protein similar to syntenin with extremely high amino acid sequence identity, when compared to the human protein, of 56% and 64%, for PDZ1 and PDZ2 domains, respectively [28]. We conducted a BLAST search of the genome of the malaria vector *Anopheles gambiae*, and found a protein annotated as a syntenin, with 50% amino acid identity to the human molecule. It is noteworthy that both the *Anopheles* and *Penaeus* homologs are far more similar to the human protein than any other PDZ domain in the human genome (Figure 6). *Ciona intestinalis*, a primitive tunicate with the smallest known genome among *Chordata*, shows the presence of sequences highly similar (50%–60% identity) to the human protein. The high sequence similarity among such diverse species suggests that the molecule predates the appearance of vertebrates.

### Syntenin: A Link between Syndecan and the Actin Cytoskeleton

The biological function of syntenin and its domain structure appear to have been stringently guarded by evolutionary mechanisms. The present study strongly supports earlier suggestions that merlin is a physiologically relevant partner for syntenin. Because merlin is an actin binding protein, syntenin may provide another link for syndecan-regulated signaling to the cytoskeleton, with syntenin meditating the colocalization of syndecan, through PDZ2, and merlin, through PDZ1. It will be comparable to the current model of syndecan signaling to actin, through PDZ-containing CASK and protein 4.1 [29] or direct binding to another of the ERM proteins, ezrin [30, 31]. The FERM domain of merlin binds ezrin and could block the interaction of ezrin to actin [31]. This alternate anchoring signal pathway may give clues regarding the involvement of syntenin in metastasis or the tumor suppressor function of merlin. The identification of the syntenin homolog in *Anopheles* prompted us to look in the mosquito's genome for the homologs of merlin and syndecan. We found an annotated merlin homolog with a 57% amino acid identity to the human protein and fully conserved C-terminal RVAFFEEL sequence. Similarly, we found the presence of a syndecan-related protein, with a highly conserved C terminus containing the TNEFYA motif.

### Biological Implications

Syntenin is a ubiquitous protein involved in protein targeting and multiprotein assembly, and it is overexpressed in certain cancer cell lines. As inferred from numerous yeast two-hybrid screens and other biochemical assays, syntenin binds biologically important receptors such as IL5Rα and syndecan, as well as the cytosolic actin regulator merlin, which is a tumor suppressor and a product of the causal gene of neurofibromatosis type II. The crystal structure of the biologically functional fragment of syntenin, residues 113–273, solved at 1.94 Å resolution, reveals the presence of two canonical PDZ domains, connected by a 4 residue linker. Both domains appear to be free to interact with target peptides. It is the first crystal structure containing more than one PDZ

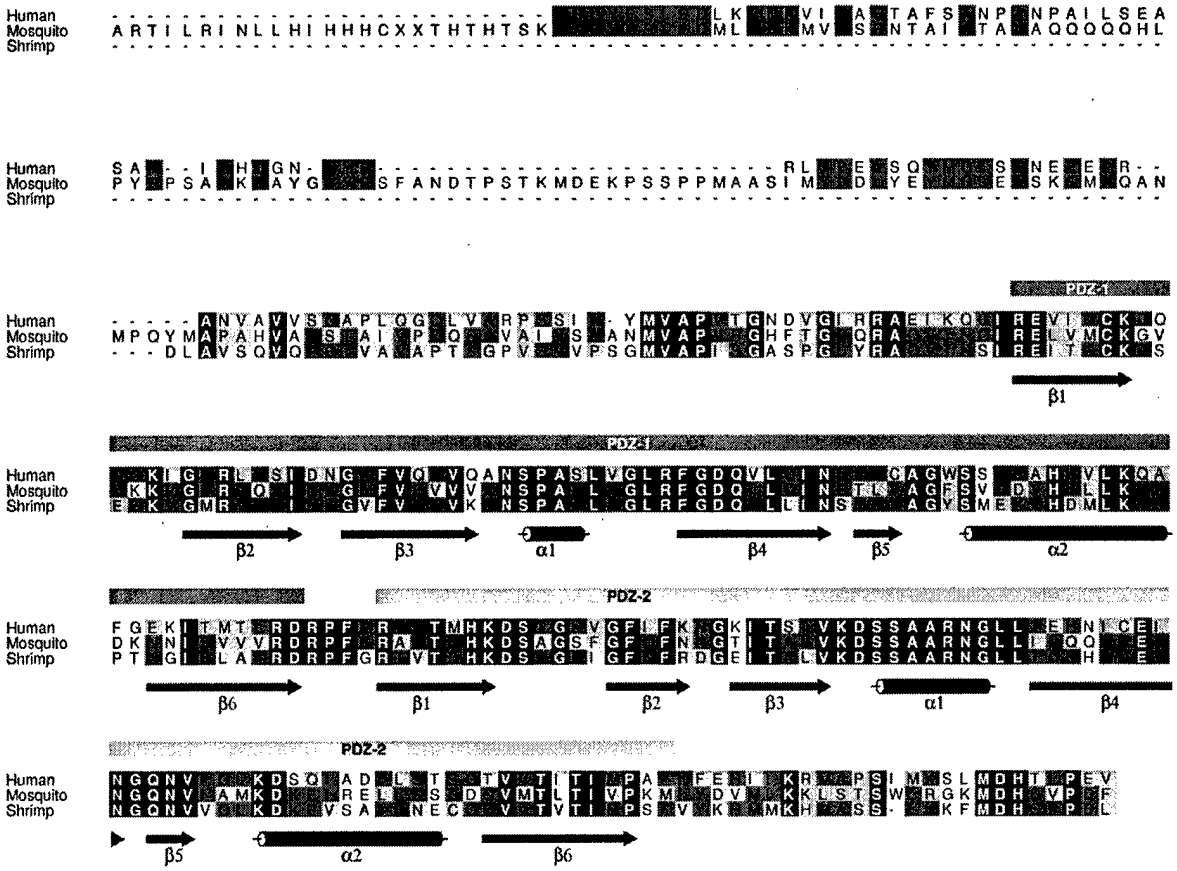



Figure 6. Amino Acid Sequence Alignment of Human Syntenin with the *Anopheles* and *Penaeus* Homologs
The secondary structural elements shown correspond to the PDZ tandem presented in this work.

domain from a single protein. Our binding studies, using stringent biophysical techniques such as isothermal titration calorimetry and fluorimetric titration, show that the properties of the tandem are the sum of the binding properties of the individual domains, with no detectable cooperative effects. Each domain is able to bind peptides belonging to two different classes: PDZ1 binds peptides corresponding to merlin and IL5Rα (classes I and III), whereas PDZ2 interacts with peptides derived from IL5Rα and syndecan-4 (classes I and II). The separate interactions of merlin with PDZ1 and that of syndecan-4 with PDZ2 suggest the physiological coupling of syndecan to merlin through syntenin. Because merlin binds actin, this pathway could be vital for merlin's function as a tumor repressor. The recently completed genome of the malaria vector *Anopheles gambiae* contains homologs of both syntenin and merlin, indicating that this pathway has been conserved during evolution.

## Experimental Procedures

### Expression and Purification of Protein Samples

A syntenin clone was obtained from American Tissue and Culture Collection (ATCC 72537). The DNA encoding full-length (residues 1–298), PDZ tandem (113–273), PDZ1 (113–193), and PDZ2 (197–273 and 197–298) domains of syntenin were amplified by PCR and cloned into the parallel vector pGST-parallel1 [32], a GST-fusion protein expression vector containing the recombinant TEV protease (rTEV) cleavage site. The integrity of the insert was verified by direct DNA sequencing. The expression of the proteins was induced by 1 mM IPTG in *E. coli* BL21 strain (Stratagene). The SeMet-labeled PDZ tandem was expressed in the D834 strain (Novagen) with M9 medium with addition of SeMet. The expressed proteins were purified by affinity chromatography using a glutathione-Sepharose 4B (Amersham Pharmacia Biotech). The recombinant protein was subjected to a HiPrep 26/10 desalting column (Amersham Pharmacia Biotech) equilibrated with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, and was digested using rTEV (Life Technologies) at 10°C in the presence of 0.5 mM EDTA, 1 mM DTT. After complete digestion, the GST tag was removed using a glutathione-Sepharose 4B column. A gel filtration was performed with a Superdex G75 column (Amersham Pharmacia Biotech) equilibrated with 20 mM Tris-HCl (pH 7.5), and the fractions containing the PDZ tandem were collected and concentrated using Centriprep YM10 for crystallization screening. The purified proteins contain an additional five amino acids (GAMDP) at the N terminus due to the cloning procedure.

### Crystallization and Data Collection

Crystal Screen (Hampton Research) was used for preliminary screening. Subsequently, crystallization conditions were optimized around 0.1 M sodium acetate (pH 4.6) containing 24% PEG4000 and 0.2 M ammonium acetate. The sitting drop vapor diffusion method was used for all crystallization trials. Drops were formed with 3 μl of protein solution (6 mg/ml) and 3 μl of reservoir buffer, and were overlaid with 15 μl of a 1:1 mixture of silicon and mineral oil. Crystallization trays were stored at 21°C. The best crystals were obtained after microseeding. The crystals used for data collection

were briefly soaked in the crystallization buffer containing 12.5% (v/v) glycerol and frozen by immersion in liquid nitrogen. The structure was solved using a three-wavelength MAD experiment with SeMet-labeled crystals, with 180 images (1° rotation) collected for each wavelength. Data were collected at 0.97946 Å (edge wavelength), 0.97133 Å (remote wavelength), and 0.97900 Å (peak wavelength). The crystals are in space group C2, with unit cell parameters a = 100.7 Å, b = 48.7 Å, c = 74.7 Å, and β = 120.8°. All data were collected at beamline X9B at the NSLS, and processed and scaled using HKL2000 [33]. Data collection statistics are presented in Table 1. The programs SnB [34] and SHELXS [35] were used to locate six of the eight selenium atoms in the asymmetric unit. Phases were generated in SHARP [36] and improved by density modification in SOLOMON [37]. These phases were used as the starting point for automatic model building in ARP/wARP [38]. This generated 275 of the 332 residues in the asymmetric unit. After manually determining which of the resulting polypeptide chains belonged to each monomer in the asymmetric unit, ARP/wARP [38] was used to dock the side chains. Manual model building was performed in O [39]. All but one of the 322 syntenin residues in the asymmetric unit were included in the model, as were 4 N-terminal residues that are an artifact of subcloning. Solvent was added using ARP/wARP [38]. A combination of CNS [40] and REFMAC5 [41] was used to refine this initial model to an R factor of 18.4% and an $R_{free}$ of 22.7%. Maximum likelihood residuals were used throughout the refinement process. TLS refinement [42] and inclusion of experimental phase information [43] from SHARP were included in later stages of refinement to minimize the difference between $R_{work}$ and $R_{free}$. MOLPROBITY [26] and OOPS2 [44] were used as validation tools during refinement and rebuilding. Refinement data are presented in Table 1.

### Calorimetric Binding Assays

Prior to the experiment, the protein solution was extensively dialyzed at 4°C against 25 mM phosphate or 25 mM Tris-HCl, 150 mM NaCl (pH 8.0). The titration was performed using a 4200 isothermal titration calorimeter (CSC). The protein concentration in the sample cell was in the range of 0.1 to 0.3 mM with a cell volume of 1,250 μl. The titrated peptides were dissolved to the concentrations in the range of 3 to 8 mM in dialysis buffer and injected in 5–10 μl aliquots. All experiments were done at 25°C. The titration thermogram was analyzed with BindWorks Applied Thermodynamics software to obtain n, $K_d$, and ΔH values. Concentration of PDZ tandem and full-length syntenin was estimated using the $A_{280}$ molar absorbance coefficient calculated from the number of Trp and Tyr residues [45]. The concentration of PDZ2 and peptides was estimated using the $A_{257}$ molar absorbance coefficient calculated from the number of Phe residues in the molecules.

### Fluorimetric Titrations

Binding of peptides to full-length syntenin and PDZ tandem and isolated domains did not produce detectable change in fluorescence. Therefore, N-terminally dansylated peptides were used to increase sensitivity. The concentration of dansylated peptide was determined using the molar absorbance coefficient of the dansyl group $\epsilon_{334} = 4{,}600\ M^{-1}cm^{-1}$. The binding was monitored by following the increase in fluorescence upon titration of a concentrated protein into a 1 cm × 1 cm stirred cell cuvette containing 1.2 ml of 25 mM Tris-HCl, 150 mM NaCl (pH 7.5) and 0.5 μM dansylated peptide. The protein stock concentration was in the range of 1–1.5 mM and the signal was corrected for the dilution factor. Data were fitted to the following equation [46] by nonlinear least squares analysis using the program Grafit 3.01 (Erithacus Software):

$$y = F_o + \frac{(F_{max} - F_o)\frac{x}{K_d}}{1 + \frac{x}{K_d}} \quad (1)$$

where y is the fluorescence signal, x is the concentration of ligand, $K_d$ is the dissociation constant, $F_o$ is the initial fluorescence value, and $F_{max}$ is the fluorescence value at saturation. Experiments were done in duplicate at 21°C using an FP-750 spectrofluorimeter (Jasco) under following conditions: λ excitation = 335 nm, λ emission = 540 nm, and excitation and emission slit width = 5 nm.

### Stability Measurements

Solvent denaturations were performed on a J-715 spectropolarimeter (Jasco) at 21°C with the automatic titrator (Jasco automatic titration system) in 25 mM Tris-HCl, 50 mM NaCl (pH 7.4) and the indicated concentration of guanidinium chloride (GdmCl). The transitions were monitored by the decrease of the CD signal at 222 nm and at 2 nm bandwidth. The apparent free-energy changes in the absence of GdmCl ($\Delta G_{un}$) were determined by fitting the ellipticity intensity changes at particular concentrations of GdmCl to the equation given elsewhere [46]. Analysis of the data was performed by the program Grafit 3.01 (Erithacus Software). GdmCl concentration was determined by refractometry.

### Acknowledgments

This work was supported by the DOD grant DAMD17-01-1-0720 (to Z.S.D.). A NATO Collaborative Link grant to Z.S.D. and J.O. is gratefully acknowledged. J.O. is an International Scholar of the Howard Hughes Medical Institute. We thank Mary Lewis for excellent technical assistance. Z.S.D. thanks Dr. R.L. Biltonen (University of Virginia) for invaluable discussions.

Received: November 26, 2002
Revised: January 14, 2003
Accepted: January 24, 2003
Published: April 1, 2003

### References

1. Grootjans, J.J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Durr, J., and David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. Proc. Natl. Acad. Sci. USA *94*, 13683–13688.
2. Fialka, I., Steinlein, P., Ahorn, H., Bock, G., Burbelo, P.D., Haberfellner, M., Lottspeich, F., Paiha, K., Pasquali, C., and Huber, L.A. (1999). Identification of syntenin as a protein of the apical early endocytic compartment in Madin-Darby canine kidney cells. J. Biol. Chem. *274*, 26233–26239.
3. Fernandez-Larrea, J., Merlos-Suarez, A., Urena, J.M., Baselga, J., and Arribas, J. (1999). A role for a PDZ protein in the early secretory pathway for the targeting of proTGF-α to the cell surface. Mol. Cell *3*, 423–433.
4. Zimmermann, P., Meerschaert, K., Reekmans, G., Leenaerts, I., Small, J.V., Vandekerckhove, J., David, G., and Gettemans, J. (2002). $PIP_2$-PDZ domain binding controls the association of syntenin with the plasma membrane. Mol. Cell *9*, 1215–1225.
5. Geijsen, N., Uings, I.J., Pals, C., Armstrong, J., McKinnon, M., Raaijmakers, J.A., Lammers, J.W., Koenderman, L., and Coffer, P.J. (2001). Cytokine-specific transcriptional regulation through an IL-5Rα interacting protein. Science *293*, 1136–1138.
6. Koo, T.H., Lee, J.J., Kim, E.M., Kim, K.W., Kim, H.D., and Lee, J.H. (2002). Syntenin is overexpressed and promotes cell migration in metastatic human breast and gastric cancer cell lines. Oncogene *21*, 4080–4088.
7. Koroll, M., Rathjen, F.G., and Volkmer, H. (2001). The neural cell recognition molecule neurofascin interacts with syntenin-1 but not with syntenin-2, both of which reveal self-associating activity. J. Biol. Chem. *276*, 10646–10654.
8. Hirbec, H., Perestenko, O., Nishimune, A., Meyer, G., Nakanishi, S., Henley, J.M., and Dev, K.K. (2002). The PDZ proteins PICK1, GRIP, and syntenin bind multiple glutamate receptor subtypes. Analysis of PDZ binding motifs. J. Biol. Chem. *277*, 15221–15224.
9. Torres, R., Firestein, B.L., Dong, H., Staudinger, J., Olson, E.N., Huganir, R.L., Bredt, D.S., Gale, N.W., and Yancopoulos, G.D. (1998). PDZ proteins bind, cluster, and synaptically colocalize with Eph receptors and their ephrin ligands. Neuron *21*, 1453–1463.
10. Lin, D., Gish, G.D., Songyang, Z., and Pawson, T. (1999). The

carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif. J. Biol. Chem. *274*, 3726–3733.
11. Iuliano, R., Trapasso, F., Sama, I., Le Pera, I., Martelli, M.L., Lembo, F., Santoro, M., Viglietto, G., Chiariotti, L., and Fusco, A. (2001). Rat protein tyrosine phosphatase η physically interacts with the PDZ domains of syntenin. FEBS Lett. *500*, 41–44.
12. Grootjans, J.J., Reekmans, G., Ceulemans, H., and David, G. (2000). Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin. J. Biol. Chem. *275*, 19933–19941.
13. Jannatipour, M., Dion, P., Khan, S., Jindal, H., Fan, X., Laganiere, J., Chishti, A.H., and Rouleau, G.A. (2001). Schwannomin isoform-1 interacts with syntenin via PDZ domains. J. Biol. Chem. *276*, 33093–33100.
14. Evans, D.G., Sainio, M., and Baser, M.E. (2000). Neurofibromatosis type 2. J. Med. Genet. *37*, 897–904.
15. den Bakker, M.A., Riegman, P.H., Suurmeijer, A.P., Vissers, C.J., Sainio, M., Carpen, O., and Zwarthoff, E.C. (2000). Evidence for a cytoskeleton attachment domain at the N-terminus of the NF2 protein. J. Neurosci. Res. *62*, 764–771.
16. Sheng, M., and Sala, C. (2001). PDZ domains and the organization of supramolecular complexes. Annu. Rev. Neurosci. *24*, 1–29.
17. Hung, A.Y., and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. J. Biol. Chem. *277*, 5699–5702.
18. Lee, S.S., Glaunsinger, B., Mantovani, F., Banks, L., and Javier, R.T. (2000). Multi-PDZ domain protein MUPP1 is a cellular target for both adenovirus E4-ORF1 and high-risk papillomavirus type 18 E6 oncoproteins. J. Virol. *74*, 9680–9693.
19. Hsueh, Y.P., Wang, T.F., Yang, F.C., and Sheng, M. (2000). Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. Nature *404*, 298–302.
20. Okamoto, M., and Sudhof, T.C. (1997). Mints, Munc18-interacting proteins in synaptic vesicle exocytosis. J. Biol. Chem. *272*, 31459–31464.
21. Dong, H., O'Brien, R.J., Fung, E.T., Lanahan, A.A., Worley, P.F., and Huganir, R.L. (1997). GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors. Nature *386*, 279–284.
22. Dong, H., Zhang, P., Song, I., Petralia, R.S., Liao, D., and Huganir, R.L. (1999). Characterization of the glutamate receptor-interacting proteins GRIP1 and GRIP2. J. Neurosci. *19*, 6930–6941.
23. Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell *85*, 1067–1076.
24. Morais Cabral, J.H., Petosa, C., Sutcliffe, M.J., Raza, S., Byron, O., Poy, F., Marfatia, S.M., Chishti, A.H., and Liddington, R.C. (1996). Crystal structure of a PDZ domain. Nature *382*, 649–652.
25. Bezprozvanny, I., and Maximov, A. (2001). Classification of PDZ domains. FEBS Lett. *509*, 457–462.
26. Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2002). Structure validation by C-α geometry: $\phi$, $\psi$ and C-β deviation. Proteins *50*, 437–450.
27. Harris, B.Z., Hillier, B.J., and Lim, W.A. (2001). Energetic determinants of internal motif recognition by PDZ domains. Biochemistry *40*, 5921–5930.
28. Bangrak, P., Graidist, P., Chotigeat, W., Supamattaya, K., and Phongdara, A. (2002). A syntenin-like protein with postsynaptic density protein (PDZ) domains produced by black tiger shrimp *Penaeus monodon* in response to white spot syndrome virus infection. Dis. Aquat. Organ. *49*, 19–25.
29. Cohen, A.R., Woods, D.F., Marfatia, S.M., Walther, Z., Chishti, A.H., Anderson, J.M., and Wood, D.F. (1998). Human CASK/LIN-2 binds syndecan-2 and protein 4.1 and localizes to the basolateral membrane of epithelial cells. J. Cell Biol. *142*, 129–138.
30. Granes, F., Urena, J.M., Rocamora, N., and Vilaro, S. (2000). Ezrin links syndecan-2 to the cytoskeleton. J. Cell Sci. *113*, 1267–1276.
31. Nguyen, R., Reczek, D., and Bretscher, A. (2001). Hierarchy of merlin and ezrin N- and C-terminal domain interactions in homo- and heterotypic associations and their relationship to binding of scaffolding proteins ebp50 and e3karp. J. Biol. Chem. *276*, 7621–7629.
32. Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of "Parallel" expression vectors. Protein Expr. Purif. *15*, 34–39.
33. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *A276*, 307–326.
34. Weeks, C.M., and Miller, R. (1999). The design and implementation of SnB v2.0. J. Appl. Crystallogr. *32*, 120–124.
35. Sheldrick, G.M., and Gould, R.O. (1995). Structure solution by iterative peaklist optimisation and tangent expansion in space group P1. Acta Crystallogr. *B51*, 423–431.
36. de la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol *276*, 472–494.
37. Abrahams, J.P., and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. *D52*, 30–42.
38. Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.
39. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110–119.
40. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905–921.
41. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. *D53*, 240–255.
42. Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. *D57*, 122–133.
43. Pannu, N.S., Murshudov, G.N., Dodson, E.J., and Read, R.J. (1998). Incorporation of prior phase information strengthens maximum-likelihood structure refinement. Acta Crystallogr. *D54*, 1285–1294.
44. Kleywegt, G.J. (2000). Validation of protein crystal structures. Acta Crystallogr. *D56*, 249–265.
45. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Sci. *4*, 2411–2423.
46. Santoro, M.M., and Bolen, D.W. (1992). A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. Biochemistry *31*, 4901–4907.
47. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
48. Merritt, E.A., and Bacon, D.J. (1997). Raster3D. Photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.
49. Nicholls, A., Sharp, K., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

## Accession Numbers

The coordinates of the syntenin PDZ tandem were deposited in the Protein Data Bank under the ID code 1N99.

Structure, Vol. 11, 845–853, July, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0969-2126(03)00125-4

# Molecular Roots of Degenerate Specificity in Syntenin's PDZ2 Domain: Reassessment of the PDZ Recognition Paradigm

Beom Sik Kang, David R. Cooper, Yancho Devedjiev, Urszula Derewenda, and Zygmunt S. Derewenda*
Department of Molecular Physiology and Biological Physics
The Cancer Center
University of Virginia
Charlottesville, Virginia 22908

## Summary

**Crystal structures of the PDZ2 domain of the scaffolding protein syntenin, both unbound and in complexes with peptides derived from C termini of IL5 receptor (α chain) and syndecan, reveal the molecular roots of syntenin's degenerate specificity. Three distinct binding sites ($S_0$, $S_{-1}$, and $S_{-2}$), with affinities for hydrophobic side chains, function in a combinatorial way: $S_{-1}$ and $S_{-2}$ act together to bind syndecan, while $S_0$ and $S_{-1}$ are involved in the binding of IL5Rα. Neither mode of interaction is consistent with the prior classification scheme, which defined the IL5Rα interaction as class I (-S/T-X-$\phi$) and the syndecan interaction as class II (-$\phi$-X-$\phi$). These results, in conjunction with other emerging structural data on PDZ domains, call for a revision of their classification and of the existing model of their mechanism.**

## Introduction

PDZ domains (postsynaptic density protein, disc large, and *zonula occludens*) occur within numerous multidomain cytosolic proteins and mediate their binding to receptors and channels, thereby serving as a membrane-associated scaffold for the assembly of signaling complexes (Harris and Lim, 2001; Hung and Sheng, 2002). Over 440 domains of this type have been identified so far in the human genome, and they are also abundant in other organisms (Sheng and Sala, 2001). PDZ domains are structurally conserved modules about 90 amino acids in size. The majority are believed to function by binding the C-terminal tail of the target protein in a structurally conserved groove between the β2 strand and the α2 helix (Doyle et al., 1996). The terminal carboxylate of the target is anchored via hydrogen bonds from three main chain amides within a conserved glycine-rich loop, a fingerprint of the PDZ fold. Early data derived from crystallographic and NMR studies suggested a general model of sequence pattern recognition, in which the peptide is bound in an extended conformation so that two side chains, $P_0$ and $P_{-2}$, point into the groove of the PDZ domain and account for specificity ($P_0$ denotes the C-terminal residue of the bound peptide and $P_{-n}$ denotes the *n*th amino acid upstream of it; $S_{-n}$ denotes the corresponding binding pocket of the PDZ domain). Those domains that are grouped together as class I bind Ser or Thr in $P_{-2}$ and a hydrophobic residue in $P_0$, so that the target sequence motif is -S/T-X-$\phi$ ($\phi$ represents hydrophobic residues and $\Psi$ represents aromatic residues). Class II domains bind another hydrophobic residue at $P_{-2}$ (-$\phi$-X-$\phi$), while a negatively charged residue at $P_{-2}$ defines class III interactions (-D/E-X-$\phi$). This simple model is unable to explain an increasing number of PDZ-mediated interactions that do not conform to this canonical type of recognition. To account for them, new classes of PDZ domains are being proposed to extend the model. For example, PDZ1 of Mint1 has been termed "novel class III" (-E/D-X-W-C/S) (Maximov et al., 1999), and PDZ3 of hINADL has been placed in "class IV" (-X-$\Psi$-D/E) (Vaccaro and Dente, 2002). These new classes of PDZ domains recognize $P_{-1}$ instead $P_{-2}$. To further complicate the issue, some PDZ domains recognize more than one class of the C-terminal sequence motif. CIPP PDZ3 binds neurexin (class II) and the NMDA receptor (class I) (Kurschner et al., 1998), and the third PDZ domain of hINADL binds the sequences -$\Psi$-D-$\phi$ (class II) and -X-$\Psi$-D sequence (class IV) (Vaccaro and Dente, 2002), while MINT1 PDZ1, hINADL PDZ5, and Par6 PDZ domains bind ligands with sequences -D-H-W-C (novel class III) and -E-Y-Y-V (class II) (Bezprozvanny and Maximov, 2001). The erbin PDZ domain binds the receptor ErbB2 (class II) and LET-23 peptide (class I) (Borg et al., 2000). While dual specificity is not rare in PDZ binding, there is no general model accounting for it.

Syntenin, first identified as a syndecan binding protein, contains a tandem of PDZ domains, which demonstrate degenerate specificity (Grootjans et al., 1997). In addition to syndecan, there are currently at least 10 binding partners reported for syntenin, including class I proteins such as interleukin 5 receptor α subunit (IL5Rα) (-D-S-V-F) (Geijsen et al., 2001), neuroglian (-Y-S-L-A) (Koroll et al., 2001), proTGF-α (-E-T-V-V) (Fernandez-Larrea et al., 1999), and neurofascin (-Y-S-L-A) (Koroll et al., 2001); class II molecules such as syndecan (-E-F-Y-A), ephrin B (-Y-Y-K-V) (Lin et al., 1999; Torres et al., 1998), Eph A7 (-G-I-Q-V) (Torres et al., 1998), PTP-η (-G-Y-I-A) (Iuliano et al., 2001), and neurexin I (-E-Y-Y-V) (Grootjans et al., 2000); and the class III protein merlin (-F-E-E-L) (Jannatipour et al., 2001). In principle, such diversity of interactions could be caused by degenerate specificity or alternatively by cooperative effects of two PDZ domains. We recently showed that syntenin's two PDZ domains show degenerate and noncooperative binding (Kang et al., 2003). The second PDZ domain (PDZ2) binds IL5Rα (class I) and syndecan-4 (class II) peptides, in spite of dramatically dissimilar sequences, with the dissociation constants of 1.9 μM and 2.3 μM, respectively. Mutational studies also show that PDZ2 has binding capacity for both class I and class II peptides (Grootjans et al., 2000; Koroll et al., 2001). In order to elucidate the molecular basis for the dual specificity of the PDZ2 domain of syntenin, we determined the crystal structures of the PDZ2 domain alone and in com-

*Correspondence: zsd4n@virginia.edu

Table 1. Crystallographic Data

| Data Set | PDZ2 | | PDZ2-Syndecan-4 | PDZ2-IL5Rα |
|---|---|---|---|---|
| Experimental Data | | | | |
| Wavelength (λ) | 1.54178 | 0.97946 | 0.97946 | 0.97946 |
| Space group | $P2_1$ | $P2_1$ | C2 | $C222_1$ |
| Unit cell parameters (Å, °) | | | | |
| a | 25.27 | 25.29 | 58.34 | 53.72 |
| b | 42.54 | 42.57 | 54.44 | 55.98 |
| c | 31.06 | 31.04 | 50.22 | 51.09 |
| β | 108.8 | 108.7 | 98.7 | 90.0 |
| Resolution (Å) | 25.0–1.60 (1.66–1.60)[a] | 13.0–1.10 (1.14–1.10) | 50.0–1.85 (1.92–1.85) | 20.0–1.25 (1.29–1.25) |
| Total reflections | 21,360 | 58,393 | 46,421 | 155,256 |
| Unique reflections | 6,600 (186) | 18,559 (275) | 12,987 (1,108) | 20,850 (1,611) |
| Completeness (%) | 79.8 (22.9) | 73.3 (11.0)[b] | 97.4 (83.7) | 95.7 (75.0) |
| $R_{sym}$ (%)[c] | 5.8 (17.0) | 4.2 (17.0) | 4.9 (41.3) | 5.1 (49.6) |
| Average $I/\sigma$ (I) | 22.7 (3.62) | 28.5 (3.46) | 25.8 (2.86) | 33.5 (2.15) |
| Refinement Details | | | | |
| Resolution (Å) | 21.27–1.60 | 12.16–1.24 | 49.39–1.85 | 19.39–1.35 |
| Reflections (working) | 6,290 | 15,685 | 12,001 | 16,171 |
| Reflections (test) | 307 | 851 | 985 | 848 |
| $R_{work}$ (%)[d] | 11.9 | 11.3 | 17.5 | 17.6 |
| $R_{free}$ (%)[d] | 16.6 | 15.3 | 22.6 | 21.2 |
| Number of waters | 181 | 173 | 145 | 144 |
| Rms deviation from ideal geometry | | | | |
| Bonds (Å) | 0.015 | 0.013 | 0.012 | 0.016 |
| Angles (°) | 1.48 | 1.63 | 1.82 | 2.13 |
| Average B factor (Å²) | | | | |
| Main chain | 11.42 | 8.32 | 16.01 | 17.19 |
| Side chain | 12.69 | 10.15 | 20.50 | 21.50 |
| Waters | 26.41 | 22.21 | 39.86 | 39.93 |

[a]The numbers in parentheses describe the relevant value for the last resolution shell.
[b]Completeness at resolution 12.16–1.24 (1.30–1.24) used for refinement is 93.1% (73.4%)
[c]$R_{sym} = \Sigma|I_i - \langle I\rangle|/\Sigma I$ where $I_i$ is the intensity of the $i$th observation and $\langle I\rangle$ is the mean intensity of the reflections.
[d]$R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma\,|F_{obs}|$, crystallographic R factor, and $R_{free} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$ when all reflections belong to a test set of randomly selected data.

plexes with an IL5Rα C-terminal peptide (ETLEDSVF) and a syndecan-4 peptide (TNEFYA). The structures were refined to 1.24 Å, 1.35 Å, and 1.85 Å resolution, respectively. These structures show how syntenin's PDZ2 can accommodate different peptides and call for a revision of the established paradigm of PDZ domain classification.

## Results and Discussion

### Syndecan Binding Involves Interaction with $Tyr_{-1}$

The crystal structure of the PDZ2 domain with a bound syndecan-4 C-terminal hexapeptide was refined at 1.85 Å resolution to a crystallographic R value of 17.5% and $R_{free}$ of 22.6% (Table 1). The structure contains a noncrystallographic dimer of PDZ2-peptide complexes in the asymmetric unit. In both complexes, the structures of the bound syndecan-4 peptides were identical within experimental error, with the average isotropic temperature factor (B factor) of 30.9 Å² and 31.4 Å², respectively. In general, the interaction between the PDZ2 and the peptide conforms to the classical model of a strand insertion between the β2 strand and α2 helix of the PDZ domain (Figure 1A). The terminal carboxylate of the peptide accepts three hydrogen bonds from the amide nitrogens of Val209, Gly210, and Phe211. There is an additional indirect interaction with the carbonyl oxygen of Gly207 through an ordered water molecule. The main chain amide of the C-terminal residue donates a hydrogen bond to the carbonyl oxygen of Phe211 in the β2 strand. The carbonyl oxygen of Phe ($P_{-2}$) interacts with the amide group of Phe213, while its amide donates a hydrogen bond to the carbonyl oxygen of the same residue. The C-terminal Ala ($P_0$) and Phe ($P_{-2}$) of the syndecan peptide interact with PDZ2 in agreement with the canonical model of class II, as exemplified by the structure of hCASK (Daniels et al., 1998). However, the methyl group of Ala ($P_0$) is much smaller than the size of the hydrophobic pocket at $S_0$, which is formed mostly by Val209, Phe211, Phe213, and Leu258 (Figure 1B). In contrast, the benzene ring of Phe ($P_{-2}$) fits well in the corresponding $S_{-2}$ pocket formed by Phe213, Asp251, and Ala255.

Interestingly, there is an additional interaction involving Tyr ($P_{-1}$), which is lodged into the $S_{-1}$ pocket cushioned by His208, Ileu212, and Val222. The aromatic ring is involved in an off-center stacking interaction with His208. Other syntenin PDZ2 binding proteins, neurexin I (class II), and neuroglian (class I) also have Tyr in the $P_{-1}$ position (Grootjans et al., 2000; Koroll et al., 2001). The equal importance of the Phe ($P_{-2}$) and Tyr ($P_{-1}$) interactions for recognition by syntenin's PDZ2 is underscored by studies showing that mutations to Ala of either of the two residues abolish binding to syntenin (Groot-

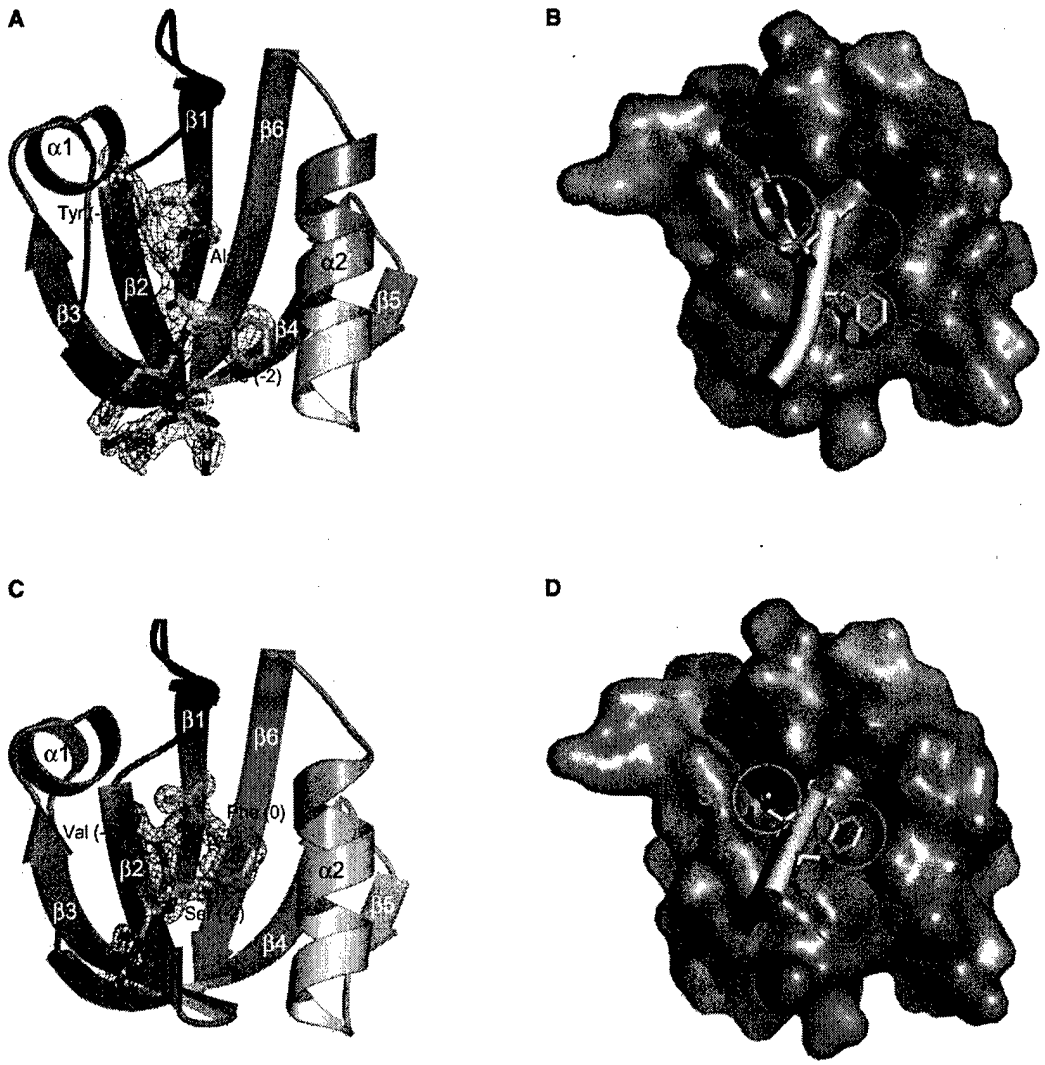



Figure 1. Comparison of Syntenin PDZ2 Structures Binding Syndecan-4 Peptide and Interleukin 5 Receptor α Subunit

(A) Ribbon diagram of the syntenin PDZ2 bound to the syndecan-4 peptide (TNEFYA). A $2mF_0 - DF_c$ electron density map calculated at 1.85 Å resolution and contoured at 1.0σ is shown around the ligand.

(B) Molecular surfaces of syntenin PDZ2 showing three hydrophobic binding pockets and the syndecan-4 peptide. The three binding pockets are circled. The three C-terminal residues are shown in the Cα trace cartoon of the peptide. The side chains of tyrosine (−1) and phenylalanine (−2) occupy the two pockets $S_{-1}$ and $S_{-2}$, while alanine (0) only occupies a portion of $S_0$.

(C) Ribbon diagram of the syntenin PDZ2 bound to the interleukin 5 receptor α subunit peptide (ETLEDSVF). A $2mF_0 - DF_c$ electron density map calculated at 1.35 Å resolution and contoured at 1.0σ is shown around the ligand.

(D) Molecular surfaces of syntenin PDZ2 showing three hydrophobic binding pockets and the interleukin 5 receptor α subunit peptide. The three binding pockets are circled. The three C-terminal residues are shown in the Cα trace cartoon of the peptide. The side chains of phenylalanine (0) and valine (−1) of the peptide are located in pockets $S_0$ and $S_{-1}$, while that of serine (−2) is out of pocket $S_{-2}$. The same orientation is used for (A) and (B) or (C) and (D). Figures were made using MOLSCRIPT (Kraulis, 1991) and Pymol (DeLano Scientific). Strand β1 (197–203); β2 (210–214); β3 (216–222); β4 (234–242); β5 (244–246); β6 (A263–270); Helix α1 (225–231); α2 (250–260).

jans et al., 1997). Thus, the interaction of PDZ2 with syndecan depends primarily on the side chains of residues in $P_{-1}$ and $P_{-2}$ rather than $P_0$ and $P_{-2}$, as the classical model implies.

### The Interaction of IL5Rα with Syntenin Does Not Involve $Ser_{-2}$

The interaction of syntenin with IL5Rα was originally reported as class I, because the C-terminal sequence of the receptor has Ser in the $P_{-2}$ position and Phe in the $P_0$ position. However, the sequences of syntenin PDZ domains do not resemble a typical class I domain. In particular, there is a notable absence of a histidine at the beginning of helix α2, which normally hydrogen bonds to the hydroxyl of Ser or Thr ($P_{-2}$). In an effort to characterize the details of the IL5Rα interactions with syntenin, we crystallized and solved the structure of the PDZ2 domain with an octapeptide derived from the

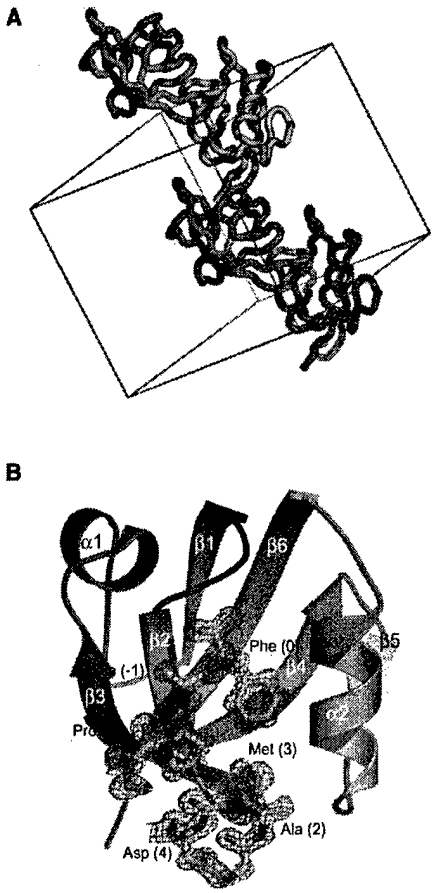



Figure 2. Syntenin PDZ2 Interaction with Its Neighbor Molecular Shows Novel PDZ Binding

(A) Crystal packing of syntenin PDZ2. Each C terminus serves as a ligand for a neighboring PDZ2 molecule.

(B) Ribbon diagram of syntenin PDZ2 bound to the C-terminal and internal sequences of its neighbor molecule. A $2mF_o - DF_c$ electron density map calculated at 1.24 Å resolution and contoured at 1.0σ around the residues bound in the PDZ domain.

(C) Molecular surfaces of syntenin PDZ2 showing three hydrophobic binding pockets and the residues of neighboring molecule binding at the pockets. The three binding pockets are circled. The three residues are shown in the Cα trace cartoon of bound residues. The side chains of phenylalanine (0) and alanine (−1) of the C terminus reside in pockets $S_0$ and $S_{-1}$, while that of methionine (3) is in pocket $S_{-2}$. The same orientation is used for (B) and (C). Figures were made using MOLSCRIPT (Kraulis, 1991) and Pymol (DeLano Scientific).

C-terminal sequence of IL5Rα. The crystals of the complex allowed for X-ray data collection to a resolution of 1.35 Å (Table 1). The atomic model, refined to a crystallographic R value of 17.6% ($R_{free}$ 21.2%), shows how the C-terminal carboxylate group of Phe ($P_0$) of the peptide is bound in a way analogous to that seen in the syndecan complex, while its benzene ring fills the $S_0$ pocket (Figures 1C and 1D). However, Ser ($P_{-2}$) does not directly interact with the PDZ2 domain as suggested by the classical model. The side chain of Val ($P_{-1}$) fits into the hydrophobic $S_{-1}$ pocket, and the carbonyl oxygen of Ser ($P_{-2}$) is hydrogen bonded through a water molecule to the Ile212 main chain nitrogen. There are no further interactions, and the electron density for the peptide fades beyond $P_{-4}$. Except for the interaction of Phe ($P_0$), the peptide's backbone does not fully occupy the binding groove as is seen in other complexes, leaving the $S_{-2}$ site empty. Similar interaction of these three C-terminal residues was also found in the crystal structure of the syntenin PDZ tandem-IL5Rα peptide complex (data not shown, our PDB entry 1OBZ). It has been shown by mutational studies that with the exception of the C-terminal Phe, no other residue is vital for the interaction of IL5Rα with syntenin (Geijsen et al., 2001). This is in agreement with our results but does not support the classical class I recognition mechanism. In contrast, the structure reveals some similarities to the conformation observed in erbin-ErbB2 peptide complex (Birrane et al., 2003). Failure of Val ($P_{-2}$) to form a typical class I interaction with His at α2 helix causes the displacement of the peptide backbone away from the α helix. In the complex of erbin and phosphorylated ErbB2 peptide (EpYLGLDVPV) complex, no density is observed beyond $P_{-5}$, leaving its other binding pocket at the β2-β3 loop empty.

### The Structure of Unbound PDZ2 Suggests Additional Recognition Mechanisms

PDZ domains show high affinity toward terminal carboxyl groups of peptides, and in the absence of target peptides, isolated PDZ domains will often bind their own C-terminal tails. The crystal structures of the hCASK PDZ, NHERF PDZ1, and GRIP1 PDZ6 domains show how the peptide binding grooves are occupied by the C-terminal tails of neighboring molecule, mimicking the recognition of the peptide ligand (Daniels et al., 1998; Im et al., 2003; Karthikeyan et al., 2001a). In a similar way, the crystal structure of the uncomplexed PDZ2 domain, refined at 1.24 Å resolution to a crystallographic R value of 11.3% ($R_{free}$ 15.3%), shows an interesting interaction between adjacent molecules (Table 1 and Figure 2A). This interaction suggests that PDZ2 is capable of yet another mode of molecular recognition, in addition to the classic mode of strand insertion.

The PDZ2 construct used here terminates at residue Phe273, rather than Met270, as is the case with the

crystal structures of the two above described complexes. (Chronologically, this structure was done first, and the construct was then truncated to circumvent inter-PDZ2 interactions and inhibition of peptide binding.) One molecule of PDZ2 binds the C-terminal Phe of its crystallographic neighbor in a manner identical to that observed for the IL5Rα (Figures 2B and 2C). The preceding residues are out of the binding groove, and there is no interaction of $P_{-2}$ at $S_{-2}$. Compared to either the syndecan or IL5Rα bound structures, the binding groove is not altered except for the side chain of Ile212, which rotates to accommodate the C terminus of the neighboring molecule. In addition to the interaction involving the C-terminal Phe, the N-terminal portion of the same molecule also interacts with the binding groove. This is possible because the C terminus and N terminus of PDZ2 are close to each other and form a structural epitope. The side chain of the third residue Met binds into the $S_{-2}$ site, so that its methyl group is in a very close contact with the aromatic ring of Phe213. A salt bridge forms between carboxyl group of an Asp, which follows the Met, and amino group of Lys214 in β2 strand of the adjacent molecule. Thus, in this case, syntenin's PDZ2 domain shows affinity for a *structural* epitope, rather than a *sequence* motif. It is very probable that this mode of recognition also occurs in nature, and that the two binding sites ($S_0$ and $S_{-2}$) may bind amino acids, which need not be adjacent within a short peptide.

### Molecular Basis of Recognition by PDZ Domains

Our results, in conjunction with existing literature, call into question the utility of the current model of protein recognition by PDZ domains and of the rigid classification of PDZ domains based on the identity of $P_0$ and $P_{-2}$ residues of a target peptide (Figure 3A). As shown here, syntenin PDZ2 has three distinct binding pockets ($S_0$, $S_{-1}$, and $S_{-2}$), and the interaction of the $P_{-1}$ residue at the $S_{-1}$ site is as important as the canonical interactions at the $S_0$ and $S_{-2}$ sites. Therefore, $P_{-1}$ interaction should be included as a general feature of PDZ interaction (Figure 3B). The importance of the $P_{-1}$ binding is also apparent from studies of other PDZ domains. Both LIN-2 and p55 PDZ domains bind peptides where all three terminal residues are hydrophobic, including aromatic side chains at both $P_{-1}$ and $P_{-2}$ (Songyang et al., 1997). By phage-displayed peptide library screening, PDZ2 of MAGI3 selects Trp as $P_{-1}$ and the side chain of this Trp is critical for high-affinity binding (Fuh et al., 2000). Model studies suggest that the side chain at $P_{-1}$ position reaches across the β2 strand and makes specific contacts with side chains in the β3 strand. The binding specificity studies of hINADL reveal that PDZ1, 2, 3, and 4 belong to class II while PDZ5, 6, and 7 are class I PDZ domains (Vaccaro et al., 2001). However, except for PDZ7, all domains in hINADL have some selectivity for $P_{-1}$, and the site-directed mutagenesis of the residues in β3 of hINADL PDZ7 alters the selectivity for $P_{-1}$. The atomic model of the NHERF PDZ domain complexed with the -Q-D-T-R-L target sequence also reveals recognition mediated by the $P_{-1}$ residue: the Arg side chain in this position interacts intimately with Asn22 and Glu43 of the PDZ (Karthikeyan et al., 2001b); these residues, lining the $S_{-1}$ pocket, are equivalent to His208 and Val222 of syntenin PDZ2.

Another example is the erbin PDZ domain, normally defined as a class I because of the His at the beginning of the α2 helix and because of the target motif -S/T-X-$\phi$. Interestingly, this domain does not bind ErbB4 (-N-T-V-V) (Borg et al., 2000), but interacts with the C termini of δ-catenin, p0071, and ARVCF, which all share the sequence -D-S-W-V (Laura et al., 2002). Furthermore, Trp was exclusively selected for $P_{-1}$ for the erbin PDZ domain by phage display peptide library screening. The NMR solution structure of the erbin PDZ domain with the phage-optimized peptide (AcTGWETWV) reveals canonical interactions of $P_0$ and $P_{-2}$ residues, as well as an additional interaction involving Trp ($P_{-1}$), whose side chain reaches across strand β2 and inserts between the side chain of Arg49 and Gln51 at the end of β3 strand (Skelton et al., 2003). Clearly, a more accurate description of the target motif for erbin would be -S/T-W-$\phi$, highlighting the preference at $P_{-1}$ for Trp.

In syntenin's PDZ2, all three binding pockets ($S_0$, $S_{-1}$, and $S_{-2}$) have an apparent affinity for hydrophobic residues (-$\phi$-$\phi$-$\phi$). These pockets function in a combinatorial way to bind peptides from different targets. This interplay of the three sites appears to endow PDZ2 with the ability to bind diverse but specific sequence motifs. The complex of syntenin's PDZ2 domain with syndecan-4 peptide, previously classified as class II interaction, involves the recognition of two penultimate residues, Tyr ($P_{-1}$) and Phe ($P_{-2}$), but not of the side chain of $P_0$ (-$\phi$-$\phi$-X) (Figures 1B and 3C). On the other hand, the interaction with the IL5Rα peptide does not involve the $P_{-2}$ side chain hydroxyl, unlike other PDZ domains that interact with class I peptides (Figure 1D). Therefore, defining the IL5Rα peptide as class I partner for syntenin is questionable. Our data suggest that syntenin's PDZ2 interactions with other so-called class I peptides may also involve binding of $P_0$ and $P_{-1}$ (-X-$\phi$-$\phi$) instead of $P_0$ and $P_{-2}$ as the classical model requires (Figure 3D). All of syntenin's partners known to bind to PDZ2 have hydrophobic residues in $P_{-1}$.

Our combinatorial model incorporates the classical $P_0$ and $P_{-2}$ interactions (-$\phi$-X-$\phi$) but accounts for all three S sites (-$\phi$-$\phi$-$\phi$) (Figures 3A and 3B). We expect that the former mode expains syntenin's interaction with ephrin B (-Y-Y-K-V), while the latter may apply to neurexin (-E-Y-Y-V). This model could also explain the dual specificity observed for other PDZ domains. For example, the CIPP PDZ3 domain binds NMDA receptors (class I) and neurexin (class II). Although the former target has a sequence -E-S-D/E-V, the CIPP PDZ3 domain does not have a His at α2 and it does not bind a related peptide from neuroligin2 (-T-T-R-V) (Kurschner et al., 1998), indicating that the $P_{-2}$ position is not critical, in contrast to canonical class I interaction. The SMART database places this particular PDZ domain in a group that binds a motif -Ψ-D/E-$\phi$ (Bezprozvanny and Maximov, 2001). Applying the combinatorial three-pocket model helps understand how it can bind NMDA receptors using $P_{-1}$ and $P_0$ (-S-D/E-V) and neurexin using $P_{-2}$ and $P_0$ (-Y-Y-V).

Aside from explicit examples of the involvement of the $P_{-1}$ residue—unaccounted for in the classical

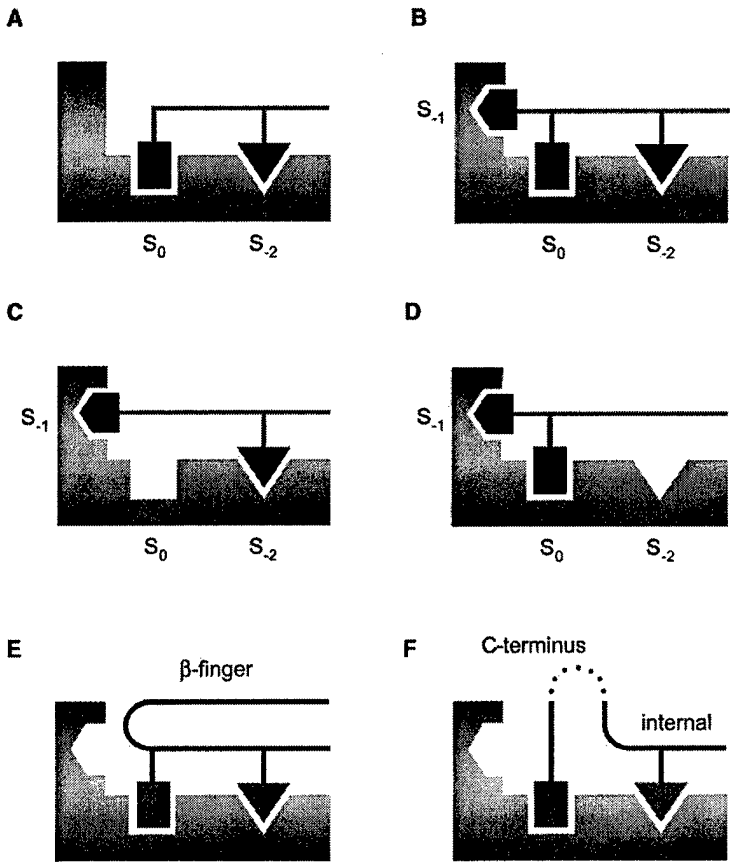



Figure 3. Schematics of PDZ Interactions

(A) Canonical PDZ binding of C-terminal sequence depends on the residues $P_0$ and $P_{-2}$ binding pockets $S_0$ and $S_{-2}$.

(B) All three C-terminal residues are involved in the interaction to PDZ binding groove.

(C) The C-terminal binding depends on the binding at $S_{-1}$ and $S_{-2}$ of PDZ as seen in PDZ2-syndecan-4 peptide complex.

(D) The C-terminal binding depends on the binding at $S_0$ and $S_{-1}$ as seen in syntenin PDZ2-interleukin 5 receptor α subunit peptide complex.

(E) Syntrophin PDZ interaction by the residues from β-finger conformation of nNOS.

(F) Interaction of internal residue at pocket $S_{-2}$ while C-terminal residues binds at $S_0$ as seen in syntenin PDZ2-PDZ2 interaction.

model—there are examples of the $P_{-2}$ residue not being involved, as in the interaction of syntenin's PDZ2 with IL5Rα. The so-called class IV or class III interactions, targeting sequences -X-Ψ-D/E or -X-W-C, do not show specificity for $P_{-2}$ (Bezprozvanny and Maximov, 2001; Maximov et al., 1999). It is not surprising that the degenerate specificity is common among these PDZ domains. Mint1 PDZ1, which has dual specificity for sequences -E/D-X-W-C/S and -E-Y-Y-V, could bind both types of sequences by interacting with $P_{-1}/P_0$ and $P_{-2}/P_{-1}$ pairs, respectively. The ability of the fifth PDZ domain of hINADL (Vaccaro et al., 2001) and PAR6 PDZ domain to bind the sequences -D-H-W-C and -E-Y-Y-V may be rationalized in a manner similar to the Mint1 PDZ1 interaction. However, there may be additional interaction involving the conserved $P_{-3}$, residue, D/E.

We believe that all three S sites described here are key determinants in the PDZ complex recognition pattern, and any general model should include all three. The binding modes of known PDZ domains, according to our combinatorial model depicted in Figure 3, are summarized in Table 2.

There is also accumulating evidence that residues upstream of the C-terminal tripeptide of the target may also be involved in the recognition process, putting the classical model in even greater peril (Laura et al., 2002; Songyang et al., 1997; Vaccaro et al., 2001). Although we found no interaction with $P_{-3}$ in the syntenin PDZ2 structures, the specificity for $P_{-3}$ is often observed. Some PDZ domains, which have long β2 strand or β2-β3 loop, have further interaction to its target peptide at this region (Birrane et al., 2003; Kozlov et al., 2002; Walma et al., 2002). We have shown that the PDZ1 domain of syntenin, which also has a long β2-β3 loop, recognizes residues upstream of the terminal hexapeptide of merlin, as exemplified by a significantly higher affinity for octathan hexapeptide (Kang et al., 2003). However, syntenin PDZ2 has a short β2-β3 loop like PSD-95 PDZ1 and NHERF PDZ1, and these domains do not interact with the bound peptides at this loop (Doyle et al., 1996; Karthikeyan et al., 2001b, 2002). Thus, the interaction including $P_{-3}$ appears to be optional, and it could enhance the binding in the absence of, or in addition to, strong binding at C-terminal three residues.

Finally, the recent erbin structures show how a binding pocket can be targeted by a residue that does not occupy the expected sequence position. The crystal structure with ErbB2 peptide shows that $Tyr_{-7}$ of the peptide binds at a site within the β2-β3 loop (Birrane et al., 2003). Interestingly, in the structure with the phage peptide (AcTGWETWV), $Trp_{-4}$ interacts at the same binding site (Skelton et al., 2003). This implies that a structural epitope is more important than the sequence for PDZ-peptide interaction. As shown by our structure of syntenin's PDZ2 and its interaction with its neighboring molecule, the residues far from the C terminus in sequence can be involved in binding to the $S_{-2}$ pocket by forming a contiguous structural epitope with the C terminus. There is also the exceptional example of the recognition of nNOS β-finger, not C terminus, by syntrophin (Figure

Table 2. Proposed Binding Modes of PDZ Domains by the Combinatorial Model

| PDZ Domains | The C-Terminal Sequence Motifs of Representive Ligands | Binding Mode[a] | References |
|---|---|---|---|
| Class I (-S/T-X-$\phi$)[b] | | | |
| PSD-95 PDZ3 | -T-S-V | A[c] | Doyle et al., 1996 |
| NHERF PDZ1 | -T-R-L, -S-L-L, -S-F-L, -E-Q-L | A[c], B[c] | Karthikeyan et al., 2001a, 2001b, 2002 |
| Erbin | -T-W-V, -V-P-V, -S-W-V | A[c], B[c] | Borg et al., 2000; Laura et al., 2002; Skelton et al., 2003 |
| hPTP1E | -S-A-V | A[c] | Kozlov et al., 2002 |
| Syntrophin | -S/T-X-V, S-L-V, T-T-F[d] | A[c], B[c] | Hillier et al., 1999; Schultz et al., 1998 |
| MAGI3 PDZ2 | -S/T-W-V | B | Fuh et al., 2000 |
| CIPP PDZ3 | -S-D/E-V, -Y-Y-V | A, C | Bezprozvanny and Maximov, 2001; Kurschner et al., 1998 |
| Class II (-$\phi$-X-$\phi$) | | | |
| hCASK | -R-E-F, -F-Y-A | A[c], C | Daniels et al., 1998 |
| Syntenin | -F-Y-A, -S-V-F, -Y-Y-V, -Y-K-V | A, B, C[c], D[c] | Geijsen et al., 2001; Grootjans et al., 1997, 2000; Torres et al., 1998; this study |
| GRIP PDZ6 | -Y-S-C | B[c] | Im et al., 2003 |
| P55 | -Y-F-I, -F-X-X, -Y-Y-F | B, C | Songyang et al., 1997 |
| LIN-2 | -F-F-V/F/A | B, C | Songyang et al., 1997 |
| Class III (-D/E-X-$\phi$) | | | |
| nNOS | -D-S-V | A[d] | Tochio et al., 1999 |
| Class IV (-X-$\Psi$-D/E, -X-W-C/S) | | | |
| hINADL PDZ3 | -W-D-V, -Y-D-W, S-W-E, -S-Y-E | B, D | Vaccaro et al., 2001 |
| Mint1 PDZ1 | -D-W-C, -H-W-C, -Y-Y-V | C, D | Bezprozvanny and Maximov, 2001; Maximov et al., 1999 |
| PAR6 | -H-W-C, -Y-Y-V | C, D | Bezprozvanny and Maximov, 2001 |
| hINADL PDZ5 | -H-W-C, -Y-Y-V, -V-F-V | C, D | Bezprozvanny and Maximov, 2001; Vaccaro et al., 2001 |

[a] A, canonical $P_0$ and $P_{-2}$ binding; B, $P_0$, $P_{-1}$, and $P_{-2}$ binding; C, $P_{-1}$ and $P_{-2}$ binding; D, $P_0$ and $P_{-1}$ binding, as shown in Figure 3.
[b] X denotes any amino acid; $\phi$ denotes a hydrophobic residue; $\Psi$ denotes an aromatic residue.
[c] Structure was solved by X-ray or NMR.
[d] $\beta$-finger binding.

3E; Hillier et al., 1999), but the mode we suggest implies involvement of internal sequences *in addition to* the interaction of the C terminus in PDZ interaction (Figure 3F).

One of the benefits of PDZ classification is the potential ability to predict the binding partners for a given PDZ domain. However, even a correctly predicted single C-terminal sequence motif may not be enough to determine the binding capacity of any given PDZ domain. For successful prediction of multiple binding partners for PDZ domains, it would be better to characterize the PDZ domain by the specificity of its binding pockets, especially the three pockets for the C-terminal residues, and account for the likely combinations.

## Biological Implications

Protein-protein interactions are pivotal to cell signaling events. The PDZ domain is the most ubiquitous protein-protein interaction module found in the human genome, with nearly 500 copies. Numerous multidomain cytosolic proteins contain PDZ domains and bind to receptors and channels, thereby serving as a membrane-associated scaffold for the assembly of signaling complexes. Syntenin is a ubiquitous protein containing two PDZ domains and is involved in protein targeting and multiprotein assembly. Notably, it is overexpressed in gastric and breast cancer cell lines, suggesting that its function contributes to cytoskeleton regulation and cell migration. Syntenin binds biologically important receptors such as IL5R and syndecan. We report the crystal structures of complexes of the second PDZ domain of syntenin, residues 113–270, with the C-terminal peptides of IL5R$\alpha$ and syndecan, solved at 1.35 Å and 1.85 Å resolution, respectively. These structures reveal how one PDZ domain interacts with different C-terminal sequences of binding targets. Syndecan binds syntenin mainly by its C-terminal $P_{-1}$ and $P_{-2}$ residues, and IL5R$\alpha$ interacts through its C-terminal $P_0$ and $P_{-1}$ residues. These results not only extend the knowledge of PDZ-ligand recognition of specific targets but also explain the general scheme underlying degenerate specificity. Furthermore, the mode of syntenin PDZ2 interaction with its neighbor molecule in a crystal of the unbound PDZ2 domain (1.24 Å resolution) suggests the importance of a structural epitope for PDZ interactions rather than a sequence motif. Based on our results and the results published elsewhere, we propose the combinatorial model that generalizes the PDZ-ligand interactions. The new model is likely to predict the possible binding of biologically important target molecules more accurately than the current model.

## Experimental Procedures

### Protein Expression and Purification

A syntenin clone was obtained from American Tissue and Culture Collection (ATCC 72537). The DNA fragment coding for PDZ2 (197–273) was amplified by PCR and cloned into a GST-fusion expression vector containing the TEV (tobacco etch virus) protease cleavage

site (Sheffield et al., 1999). This construct was used for expression of protein samples for the crystallization of uncomplexed PDZ2. To obtain a shorter version of PDZ2 (197–270), a stop codon (TGA) was introduced after Met270 by the QuikChange® method (Stratagene). This step was necessary to prevent intermolecular interactions between the PDZ2 domains, in which one molecule bound another via the terminal Phe (see text). Both versions of the PDZ2 domain were expressed in *E. coli* BL21 strain (Stratagene) and purified using glutathione-Sepharose 4B column (Amersham Pharmacia Biotech). The eluted recombinant protein was subjected to a HiPrep 26/10 desalting column (Amersham Pharmacia Biotech) equilibrated with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl (buffer A) to remove glutathione. After complete digestion with rTEV protease (Life Technologies) at 10°C in the presence of 0.5 mM EDTA and 1 mM DTT, the protein solution was passed again through a glutathione Sepharose 4B column and the flow-through was concentrated and loaded on a Superdex G75 column (Amersham Pharmacia Biotech) equilibrated with buffer A. The protein fractions were collected and concentrated. All the purification steps, except the rTEV digestion, were performed at 4°C. The purified PDZ2 domain contains an additional pentapeptide (GAMDP) at the N terminus due to the cloning procedure.

### Crystallization

Initial search for crystals of PDZ2 (197–273) was carried out with Crystal Screen 1 (Hampton Research, Inc.) using the hanging drop vapor-diffusion technique at 294K. The best crystals of PDZ2 were obtained with 8 mg/ml protein concentration at 0.1 M HEPES (pH 7.0), 34% PEG4000 using the sitting drop vapor-diffusion method with microseeding. For the crystallization of complexes of short PDZ2 (197–270) with peptides, we used Additive screen I (Hampton Research, Inc.) for additional screening. The complex of PDZ2 and syndecan-4 peptide was crystallized from 0.1 M HEPES (pH 6.8), 1.6 M ammonium sulfate, 20 mM $CoCl_2$, and 0.2 M $MgSO_4$, using 1:2 molar mixtures of protein and peptide. The best crystal of PDZ2 with the IL5Rα peptide was obtained from 0.1 M HEPES (pH 7.0), 1.6 M ammonium sulfate, 20 mM $CoCl_2$, and 0.2 M $MgSO_4$ by microseeding. Synthesized octapeptide of IL5Rα (ETLEDSVF) and hexapeptide of syndecan-4 (TNEFYA) were purchased from Bio-Synthesis and UVA Biomolecular Research Facility, respectively.

### Data Collection, Structure Determination, and Refinement

Crystals were frozen in liquid nitrogen. Those containing peptides were briefly soaked in the crystallization buffer containing 17.5% (v/v) glycerol and peptide before freezing. An initial data set was collected with an R-Axis IV detector and a Nonius FR591 generator equipped with Osmic confocal mirrors. Subsequent data were collected at beamline X9B at NSLS with a wavelength of 0.97946 Å under cryoconditions using an ADSC Quantum4 CCD. Data sets were processed and scaled using HKL2000 (Otwinowski and Minor, 1997). Crystallographic details including unit cells and data statistics are shown in Table 1. All structures were solved by the molecular replacement method using AMORE (Navaza, 1994) and the atomic models were refined with REFMAC5 (Murshudov et al., 1997) from the CCP4 suite (CCP4, 1994). The atomic coordinates of the PDZ2 domain derived from the structure of the PDZ tandem (entry 1N99 in the PDB) were used as initial model for molecular replacement of the home source data set of PDZ2. The PDZ2 structure refined to 1.60 Å was subsequently used as a model for molecular replacement of the other data sets collected using synchrotron radiation. Manual model building was performed in O (Jones et al., 1991). The final models agree well with known protein geometry. Details of refinement are given in Table 1. Compared to unbound PDZ2 in the PDZ tandem structure, the rms differences for Cα atoms of PDZ2 structures with syndecan, IL5Rα peptide, and alone but interacting with the neighboring molecule, are 0.39 Å, 0.44 Å, and 0.39 Å, respectively, indicating that the bound peptide causes no significant structural changes in the PDZ domain.

### Acknowledgments

Supported by the DOD grant DAMD17-01-1-0720 (to Z.S.D.). We thank Mary Lewis for excellent technical assistance and Dr. Zbigniew Dauter (NCI/NSLS) for help with data collection.

Received: February 13, 2003
Revised: March 3, 2003
Accepted: March 27, 2003
Published: July 1, 2003

### References

Bezprozvanny, I., and Maximov, A. (2001). Classification of PDZ domains. FEBS Lett. *509*, 457–462.

Birrane, G., Chung, J., and Ladias, J.A. (2003). Novel mode of ligand recognition by the erbin PDZ domain. J. Biol. Chem. *278*, 1399–1402.

Borg, J.P., Marchetto, S., Le Bivic, A., Ollendorff, V., Jaulin-Bastard, F., Saito, H., Fournier, E., Adelaide, J., Margolis, B., and Birnbaum, D. (2000). ERBIN: a basolateral PDZ protein that interacts with the mammalian ERBB2/HER2 receptor. Nat. Cell Biol. *2*, 407–414.

CCP4 (Collaborative Computational Project Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.

Daniels, D.L., Cohen, A.R., Anderson, J.M., and Brunger, A.T. (1998). Crystal structure of the hCASK PDZ domain reveals the structural basis of class II PDZ domain target recognition. Nat. Struct. Biol. *5*, 317–325.

Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell *85*, 1067–1076.

Fernandez-Larrea, J., Merlos-Suarez, A., Urena, J.M., Baselga, J., and Arribas, J. (1999). A role for a PDZ protein in the early secretory pathway for the targeting of proTGF-α to the cell surface. Mol. Cell *3*, 423–433.

Fuh, G., Pisabarro, M.T., Li, Y., Quan, C., Lasky, L.A., and Sidhu, S.S. (2000). Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. J. Biol. Chem. *275*, 21486–21491.

Geijsen, N., Uings, I.J., Pals, C., Armstrong, J., McKinnon, M., Raaijmakers, J.A., Lammers, J.W., Koenderman, L., and Coffer, P.J. (2001). Cytokine-specific transcriptional regulation through an IL-5Rα interacting protein. Science *293*, 1136–1138.

Grootjans, J.J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Durr, J., and David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. Proc. Natl. Acad. Sci. USA *94*, 13683–13688.

Grootjans, J.J., Reekmans, G., Ceulemans, H., and David, G. (2000). Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin. J. Biol. Chem. *275*, 19933–19941.

Harris, B.Z., and Lim, W.A. (2001). Mechanism and role of PDZ domains in signaling complex assembly. J. Cell Sci. *114*, 3219–3231.

Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S., and Lim, W.A. (1999). Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science *284*, 812–815.

Hung, A.Y., and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. J. Biol. Chem. *277*, 5699–5702.

Im, Y.J., Park, S.H., Pho, S.H., Lee, J.H., Kang, G.B., Sheng, M., Kim, E., and Eom, S.H. (2003). Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization. J. Biol. Chem. *278*, 8501–8507.

Iuliano, R., Trapasso, F., Sama, I., Le Pera, I., Martelli, M.L., Lembo, F., Santoro, M., Viglietto, G., Chiariotti, L., and Fusco, A. (2001). Rat protein tyrosine phosphatase η physically interacts with the PDZ domains of syntenin. FEBS Lett. *500*, 41–44.

Jannatipour, M., Dion, P., Khan, S., Jindal, H., Fan, X., Laganiere, J., Chishti, A.H., and Rouleau, G.A. (2001). Schwannomin isoform-1 interacts with syntenin via PDZ domains. J. Biol. Chem. *276*, 33093–33100.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density

maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kang, B.S., Cooper, D.R., Jelen, F., Devedjiev, Y., Derewenda, U., Dauter, Z., Otlewski, J., and Derewenda, Z.S. (2003). PDZ-tandem of human syntenin: crystal structure and functional properties. Structure *11*, 459–468.

Karthikeyan, S., Leung, T., Birrane, G., Webster, G., and Ladias, J.A. (2001a). Crystal structure of the PDZ1 domain of human $Na^+/H^+$ exchanger regulatory factor provides insights into the mechanism of carboxyl-terminal leucine recognition by class I PDZ domains. J. Mol. Biol. *308*, 963–973.

Karthikeyan, S., Leung, T., and Ladias, J.A. (2001b). Structural basis of the $Na^+/H^+$ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator. J. Biol. Chem. *276*, 19683–19686.

Karthikeyan, S., Leung, T., and Ladias, J.A. (2002). Structural determinants of the $Na^+/H^+$ exchanger regulatory factor interaction with the $\beta_2$ adrenergic and platelet-derived growth factor receptors. J. Biol. Chem. *277*, 18973–18978.

Koroll, M., Rathjen, F.G., and Volkmer, H. (2001). The neural cell recognition molecule neurofascin interacts with syntenin-1 but not with syntenin-2, both of which reveal self-associating activity. J. Biol. Chem. *276*, 10646–10654.

Kozlov, G., Banville, D., Gehring, K., and Ekiel, I. (2002). Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the β2-β3 loop to PDZ domain-ligand interactions. J. Mol. Biol. *320*, 813–820.

Kraulis, P.J. (1991). MOLSCRIP: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Kurschner, C., Mermelstein, P.G., Holden, W.T., and Surmeier, D.J. (1998). CIPP, a novel multivalent PDZ domain protein, selectively interacts with Kir4.0 family members, NMDA receptor subunits, neurexins, and neuroligins. Mol. Cell. Neurosci. *11*, 161–172.

Laura, R.P., Witt, A.S., Held, H.A., Gerstner, R., Deshayes, K., Koehler, M.F., Kosik, K.S., Sidhu, S.S., and Lasky, L.A. (2002). The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of delta-catenin and ARVCF. J. Biol. Chem. *277*, 12906–12914.

Lin, D., Gish, G.D., Songyang, Z., and Pawson, T. (1999). The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif. J. Biol. Chem. *274*, 3726–3733.

Maximov, A., Sudhof, T.C., and Bezprozvanny, I. (1999). Association of neuronal calcium channels with modular adaptor proteins. J. Biol. Chem. *274*, 24453–24456.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D *53*, 240–255.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Methods in Enzymology, Volume 276, C.W. Carter, Jr., and R.M. Sweet, eds. (New York: Academic Press), pp. 307–326.

Schultz, J., Hoffmuller, U., Krause, G., Ashurst, J., Macias, M.J., Schmieder, P., Schneider-Mergener, J., and Oschkinat, H. (1998). Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. Nat. Struct. Biol. *5*, 19–24.

Sheffield, P., Garrard, S., and Derewenda, Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. Protein Expr. Purif. *15*, 34–39.

Sheng, M., and Sala, C. (2001). PDZ domains and the organization of supramolecular complexes. Annu. Rev. Neurosci. *24*, 1–29.

Skelton, N.J., Koehler, M.F., Zobel, K., Wong, W.L., Yeh, S., Pisabarro, M.T., Yin, J.P., Lasky, L.A., and Sidhu, S.S. (2003). Origins of PDZ domain ligand specificity: structure determination and mutagenesis of the Erbin PDZ domain. J. Biol. Chem. *278*, 7645–7654.

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M., and Cantley, L.C. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science *275*, 73–77.

Tochio, H., Zhang, Q., Mandal, P., Li, M., and Zhang, M. (1999). Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide. Nat. Struct. Biol. *6*, 417–421.

Torres, R., Firestein, B.L., Dong, H., Staudinger, J., Olson, E.N., Huganir, R.L., Bredt, D.S., Gale, N.W., and Yancopoulos, G.D. (1998). PDZ proteins bind, cluster, and synaptically colocalize with Eph receptors and their ephrin ligands. Neuron *21*, 1453–1463.

Vaccaro, P., and Dente, L. (2002). PDZ domains: troubles in classification. FEBS Lett. *512*, 345–349.

Vaccaro, P., Brannetti, B., Montecchi-Palazzi, L., Philipp, S., Citterich, M.H., Cesareni, G., and Dente, L. (2001). Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from *Drosophila melanogaster*. J. Biol. Chem. *276*, 42122–42130.

Walma, T., Spronk, C.A., Tessari, M., Aelen, J., Schepens, J., Hendriks, W., and Vuister, G.W. (2002). Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL. J. Mol. Biol. *316*, 1101–1110.

### Accession Numbers

Coordinates have been deposited with ID codes 1NTE, 1OBY, and 1OBX for the syntenin PDZ2 alone, PDZ2-syndecan-4 peptide complex, and PDZ2-IL5Rα peptide complex, respectively.

doi:10.1016/j.jmb.2004.02.057 J. Mol. Biol. (2004) 338, 483–493

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# The PDZ2 Domain of Syntenin at Ultra-high Resolution: Bridging the Gap Between Macromolecular and Small Molecule Crystallography

**Beom Sik Kang, Yancho Devedjiev, Urszula Derewenda and Zygmunt S. Derewenda***

*Department of Molecular Physiology and Biological Physics, University of Virginia Charlottesville, VA 22908-0736 USA*

*Corresponding author

The crystal structure of the second PDZ domain of the scaffolding protein syntenin was solved using data extending to 0.73 Å resolution. The crystallographic model, including the hydrogen atoms and the anisotropic displacement parameters, was refined to a conventional *R*-factor of 7.5% and $R_{free}$ of 8.7%, making it the most precise crystallographic model of a protein molecule to date. The model reveals discrete disorder in several places in the molecule, and significant plasticity of the peptide bond, with some ω angles deviating by nearly 20° from planarity. Most hydrogen atoms are easily identifiable in the electron density and weak hydrogen bonds of the C–H···O type are clearly visible between the β-strands. The study sets a new standard for high-resolution protein crystallography.



*Keywords:* PDZ; syntenin; ultra-high resolution; crystallography

## Introduction

For several decades after the pioneering diffraction experiments with wet pepsin crystals in 1934 by J. D. Bernal and D. Crowfoot,[1] crystallographers studying macromolecules maintained firmly that protein crystals diffract weakly to a limited resolution, with only few exceptions of very small proteins. This problem was attributed to such factors as intrinsic disorder in protein crystals, large volumes filled with liquid solvent and the size of the unit cell. The advent of synchrotron radiation[2] changed this perception to some degree, but the advantage of high flux was typically applied to improve the data for poorly diffracting crystals, rather than to push the experimental envelope for crystals that were of good quality to begin with. Obtaining data to a resolution within 1.5–2.0 Å was critical, as it allowed for the application of emerging refinement methods.[3] It was noted, however, that a number of protein crystals yielded diffraction to an unexpectedly high resolution, even though other technical difficulties precluded full utilization of that potential. For example, the tetragonal crystals of the *Bacillus cereus* phospholipase C gave very strong Bragg reflections beyond 1.0 Å resolution, observed on still photographs recorded at beamline 9.6 in Daresbury (Z.S.D., unpublished results). Unfortunately, it was neither practical nor feasible in the 1980s to collect atomic-resolution data on film, using a rotation camera, and at 4 °C, so a compromise limit of 1.5 Å was used in that case instead.[4] In subsequent years, the introduction of digitally read imaging plates eventually permitted fine-slicing and adequate spatial resolution of reflections at high diffraction angles, while cryocrystallography[5] dispensed with the need to merge data from a number of crystals, which routinely died from radiation damage after a short period in the beam at temperatures above 0 °C. Consequently, it became possible to collect diffraction data to the true resolution limit, and it quickly became apparent that for a number of protein crystals this extends to atomic resolution, defined as 1.2 Å, or significantly further.[6–8]

The field of ultra-high-resolution protein crystallography is of paramount importance to structural biology, even though atomic-resolution protein structures are these days upstaged, as judged by the covers of select scientific periodicals, by low-resolution membrane protein structures, because of the immediate biomedical impact. One should

Abbreviations used: PEG, polyethylene glycol; LHB, lost hydrogen bond.

E-mail address of the corresponding author: zsd4n@virginia.edu

not forget that detailed structural information about molecules perceived as less appealing, such as crambin,[9] was vital for the development of refinement and validation techniques that lend credibility to the low-resolution models.[10]

In the course of our studies of the scaffolding protein syntenin, we have identified a crystal form of the second PDZ domain of this protein, referred to as synPDZ2, which diffracted beyond 0.8 Å resolution at the X8C beamline at NSLS (Brookhaven National Laboratory, Upton, New York). PDZ domains are structurally conserved modules about 90 amino acid residues in size.[11] Most function by binding the C-terminal tail of the target protein in a structurally conserved groove between the β2 strand and the α2 helix. The terminal carboxylate group of the target is anchored *via* hydrogen bonds from three main chain amide groups within a conserved glycine-rich loop preceding β2, a fingerprint of the PDZ fold.

For technical reasons we collected data to 0.73 Å, with completeness falling gradually beyond 0.8 Å, due to the square shape of the detector (see Materials and Methods). In spite of this deficiency, the data still compare very favorably with other reported ultra-high-resolution protein studies (see Table 1), consequently allowing for an unbiased evaluation of various aspects of crystallographic refinement and protein chemistry. The final atomic model was refined to an *R*-factor of 7.5% ($R_{free}$ 8.7%), making it the most precisely refined crystallographic model of a protein molecule to date. The model conforms very well to the expected stereochemical parameters, and highlights two important aspects: the flexibility of the peptide bond, and the distinct stereochemistry of weak, C–H···O hydrogen bonds.

## Results

### Improvement in crystal quality

We have reported the structure of synPDZ2 at 1.24 Å resolution.[12] In the present study, we were able to extend the data to the ultra-high-resolution range owing to further improvements in the quality of the crystals (Table 2). The high-quality diffraction of these crystals is probably due, in part, to compact packing, in agreement with recent statistical analyses.[13] The packing is mediated by extensive interactions between neighboring molecules along the *b*-axis. The Matthews coefficient is 1.79 Å/Da, and the solvent content is 31.2% (v/v). The unit cell volume of the new crystals is smaller by 1.7% compared to our previous study. The observed "compression" is primarily along the *b*-axis, and is associated with some subtle repacking of residues in the crystal contact region. In addition, in the ultra-high-resolution structure, we did not observe oxidation of Cys239, located near the C-terminal binding motif, as seen in the 1.24 Å resolution structure. It is not clear if this has had an effect on crystal quality. Although we used the same crystallization conditions for both experiments, we cannot rule out the possibility that the observed compressed packing is due to small differences in the concentration of polyethylene glycol (PEG) used.

**Table 1.** The highest-resolution protein structures in the PDB and comparison to synPDZ2

| Protein | PDB entry | Residues | $V_M$ (Å³/Da) | Maximum resolution (Å) | $R_{merge}$[a] (shell)[b] | Completeness (outer shell)[c] | $R$[c] | $R_{free}$[c] |
|---|---|---|---|---|---|---|---|---|
| Crambin | 1EJG | 46 | 1.40 | 22.4–0.54 | 5.5 (14.8) | 97.6 (100) | 9.0 | 9.4 |
| Antifreeze protein RD1 | 1UCS | 64 | 1.40 | 22.3–0.62 | 7.2 (65.3) | 92.8 (91.8) | 13.3 | 15.5 |
| Subtilisin (serine protease) | 1JEA | 269 | 2.31 | 35.0–0.78 | 3.6 (29.0) | 97.3 (92.7) | 9.9 | 10.3 |
| High potential iron–sulfur protein | 1IUA | 83 | NA[d] | 20.0–0.80 | 5.2 (39.7) | 98.6 (95.5) | 10.1 | 11.4 |
| Trypsin[e] | 1FY4 | 227 | NA | 20.0–0.81 | 3.8 (26.9) | 92.8 (88.4) | 10.8 | NA |
| Photoactive yellow protein | 1NWZ | 125 | 1.38 | 30.0–0.82 | NA | 97.5 (85.7) | 12.3 | 14.4 |
| Triosephosphate isomerase | 1N55 | 251 | 1.77 | 25.0–0.83 | 2.9 (39.4) | 99.3 (97.2) | NA | 10.8 |
| β-Lactamase Tem-1 | 1M40 | 263 | 1.69 | 15.0–0.85 | 5.2 (42.0) | 100 (100) | 9.1 | 11.2 |
| Acutohaemolysin | 1MC2 | 122 | 1.45 | 10.0–0.85 | 3.0 (30.2)[f] | 79.9 (30.0) | 9.5 | 12.1 |
| Trypsin inhibitor | 1G6X | 58 | 2.28 | 10.0–0.86 | 3.6 (48.8)[f] | 94.9 (8.6) | 10.7 | 14.0 |
| Xylose isomerase | 1MUW | 386 | 1.97 | 50.0–0.86 | 5.1 (56.5) | 97.4 (92.0) | 12.5 | 14.3 |
| Syntenin PDZ2 | 1R6J | 82 | 1.79 | 20.0–0.73 | 6.0 (38.7)[f] | 83.9 (9.6) | 7.5 | 8.7 |

The Table includes protein structures with higher than 0.86 Å resolution except peptide structures, α-conotoxin Si (14 residues) designed peptide α-1 (26 residues) and gramicidin D (36 residues).

[a] $R_{merge} = \sum|I_i - \langle I\rangle|/\sum I$ where $I_i$ is the intensity of the *i*th observation and $\langle I\rangle$ is the mean intensity of the reflections.

[b] The numbers in parentheses describe the relevant value for the highest-resolution shell.

[c] $R = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$, crystallographic *R* factor, and $R_{free} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$ when all reflections belong to a test set of randomly selected data.

[d] Not available.

[e] The structure of lowest *R* factor among four 0.81 Å resolution trypsin structures.

[f] Due to incompleteness of high-resolution data, the nominal resolution of these studies should be somewhat lower than the maximum resolution.

**Table 2.** Crystallographic data

| | | |
|---|---|---|
| *A. Experimental data* | | |
| Wavelength (Å) | 0.8500 | |
| Space group | $P2_1$ | |
| Unit cell parameters | | |
| *a* (Å) | 25.88 | |
| *b* (Å) | 39.54 | |
| *c* (Å) | 32.28 | |
| β (deg.) | 109.6 | |
| Resolution (Å) | 20.0–0.73 (0.76–0.73)[a] | |
| Total reflections | 316,159 | |
| Unique reflections | 71,347 (813) | |
| Completeness (%) | 83.9 (9.6) | |
| $R_{merge}$[b] (%) | 6.0 (38.7) | |
| Average $I/\sigma(I)$ | 25.3 (1.7) | |
| Wilson *B*-factor (Å$^2$) | 3.75 | |
| *B. Refinement details* | *Refmac5* | *Shelx-97* |
| Resolution (Å) | 16.6–1.0 | 10.0–0.73 |
| Reflections (working) | 33,177 | 69,900 |
| Reflections (test) | 655 | 1,389 |
| $R_{work}$[c] | | |
| All data (%) | 7.89 | 7.45 |
| $F_o > 4\sigma$ (%) | – | 6.62 |
| $R_{free}$[c] | | |
| All (%) | 9.63 | 8.66 |
| $F_o > 4\sigma$ (%) | – | 7.79 |
| Number of water molecules | 209 | 237 |
| R.M.S. deviation from ideal geometry | | |
| Bonds (Å) | 0.012 | 0.019 |
| Angles (deg.) | 1.724 | – |
| Angle bonded distances (Å) | – | 0.033 |
| Average *B*-factor | | |
| Main chain (Å$^2$) | 3.32 | 4.70 |
| Side-chain (Å$^2$) | 4.53 | 6.00 |
| Water (Å$^2$) | 10.73 | 10.82 |

[a] The numbers in parentheses describe the relevant value for the last resolution shell.

[b] $R_{merge} = \sum|I_i - \langle I\rangle|/\sum I$ where $I_i$ is the intensity of the ith observation and $\langle I\rangle$ is the mean intensity of the reflections.

[c] $R_{work} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$, crystallographic *R* factor, and $R_{free} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$ when all reflections belong to a test set of randomly selected data.

## Precision of the model refinement

Due to the square shape of the detector and the data collection geometry, we were able to record a nearly complete dataset to 0.8 Å resolution, and a progressively decreasing fraction of data to 0.73 Å (Table 2). The alternative would have been to use 2θ geometry, which was technically not possible at the time. Although only 56% complete, the data beyond 0.8 Å contain 7413 useful reflections (10% of the recorded data) and we decided to use all those data in subsequent calculations. An atomic model of synPDZ2 with isotropic displacement parameters, but without hydrogen atoms, was refined to a value for $R_{work}$ of 14.8% and $R_{free}$ of 15.7%. Anisotropic refinement decreased the $R_{work}$ to 10.8%, and $R_{free}$ to 12.5%. This is in agreement with the previously reported observations that anisotropic displacement parameters typically account for 4–5% of the *R*-factor. Adding hydrogen atoms at this stage reduced the *R*-factor to 9.61% and $R_{free}$ to 10.98%.

At this point in the refinement, we discovered that the N terminus and the β1–β2 loop have minor secondary conformations. The adjustment of the model to include this disorder had a significant impact on the agreement factors ($R_{work}$ = 8.30%, $R_{free}$ = 9.96%). Further refinement of the solvent structure and adjusting occupancies of the disordered parts and water molecules resulted in a drop of $R_{work}$ to 7.53%, and $R_{free}$ to 8.71%. The only remaining option was to introduce limited freedom for hydrogen atoms, and this proved to be marginally helpful, bringing $R_{work}$ to 7.47% and $R_{free}$ to 8.66%. At this point, we have discontinued using the $R_{free}$ and we used all the data for the final phase of calculations. Unrestrained refinement had no noticeable effect on *R* (7.45%), but restraints for minor alternative conformations had to be in place, as without these restraints the atomic coordinates of atoms with low occupancy shifted out of the density. A full-matrix refinement resulted in no change to atomic coordinates, the *R*-factor or the electron density map.

The *R*-values for the synPDZ2 model are the lowest for any protein structure recorded to date (Table 1). With the exception of crambin at 0.53 Å resolution, no prior study ever succeeded in bringing the $R_{free}$ below 10%, while our model shows 8.7%. The $R_{work}$ value is several percentage points lower than representative structures reported in the 0.78–0.85 Å resolution range.

## Multiple main chain and side-chain conformations

The N terminus of synPDZ2, Gly192 to Thr198 using the numbering of full-length syntenin, shows distinct static disorder (Figure 1). The occupancy of the minor conformer is 0.18, and the electron density for the respective atoms is only a

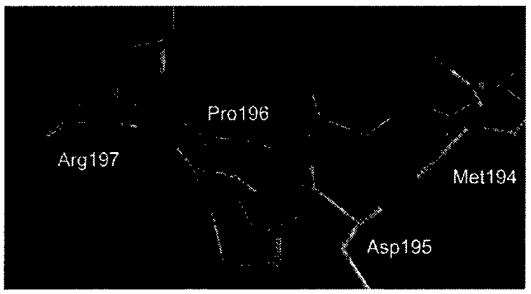



**Figure 1.** Electron density maps around the disordered N terminus of synPDZ2. The $2F_o - F_c$ electron density map (blue) and $F_o - F_c$ difference electron density map (red) from the structure with major conformer alone are contoured at +4.0 σ. Major and minor conformers of residues Met194 to Arg197 are shown as thick and thin frame, respectively. The Figures were generated using O.[20]

little higher than that of hydrogen atoms of main chain amide groups. The N-terminal amino group of the major conformer donates a hydrogen bond to the carbonyl oxygen atom of Arg229 of the neighboring molecule, while the carbonyl group of Gly192 in the minor conformer accepts a hydrogen bond from the amide group of Ser252 of another partner molecule across crystal contacts. Another apparent difference is the presence of a chloride ion near the N terminus of the major conformer. It is possible that the conformational heterogeneity is related to the partial occupancy of the chloride ion.

Another fragment of the structure with two alternative main-chain conformations is the β1–β2 loop, i.e. Asp204 to Gly207. Although the ratio of occupancies of the main to minor conformers (0.81–0.19) is similar to what is observed at the N terminus, it is unlikely that the two are interdependent, since there is no direct contact between them. This loop does not interact with any of the neighboring molecules, and the disorder is more likely to be associated with the heterogeneity of the hydrogen bonds within the loop. Hydrogen bonds between $O^{\gamma 1}$ of Thr206 and $O^{\delta 1}$ of Asp204 (2.67 Å), and $N^{\delta 1}$ of His208 and $O^{\gamma 1}$ of Thr206 (2.70 Å), are found in the major conformation but not in the minor conformation. Interestingly, the position of the loop in the previously reported PDZ2 structure is between the two high-resolution conformations, and the electron density in this region is not clear in other structures of syntenin either. It is possible that ultra-high-resolution data were able to resolve static disorder that was unidentifiable earlier.

There are alternative conformations for the side-chains of Asp195, Ser205 (both within the major main chain conformer), His202, Lys214, Ile218, Glu235, His236, Glu240, Asn245, Ser259, Thr268, and Met270. A higher number of multiple occupancies is expected with improvement in resolution.[14]

Overall, when the model is compared to the 1.24 Å resolution synPDZ2 structure, it shows an r.m.s. difference for the main-chain atoms of 0.745 Å for all residues, but the value is only 0.364 Å when the N-terminal five residues and C-terminal two residues are excluded. These residues participate in the stacking along the *b*-axis, where the unit cell compression resulted in packing rearrangements.

**Table 3.** Analysis of main chain bond lengths and angles

| | Residues | Min value | Max value | Mean value | Mean value (small molecule data)[a] |
|---|---|---|---|---|---|
| *A. Bond* | | | | | |
| C–N | Except Pro | 1.304 | 1.358 | 1.333 (0.009)[b] | 1.329 (0.014) |
| | Pro | 1.325 | 1.341 | 1.333 (0.008) | 1.341 (0.016) |
| C–O | | 1.215 | 1.267 | 1.235 (0.010) | 1.231 (0.020) |
| $C^{\alpha}$–C | Except Gly | 1.492 | 1.544 | 1.525 (0.010) | 1.525 (0.021) |
| | Gly | 1.508 | 1.543 | 1.520 (0.011) | 1.516 (0.018) |
| $C^{\alpha}$–$C^{\beta}$ | Ala | 1.510 | 1.526 | 1.519 (0.007) | 1.521 (0.033) |
| | Ile, Thr, Val | 1.520 | 1.587 | 1.540 (0.013) | 1.540 (0.027) |
| | The rest | 1.506 | 1.563 | 1.532 (0.013) | 1.530 (0.020) |
| N–$C^{\alpha}$ | Except Gly, Pro | 1.434 | 1.466 | 1.435 (0.008) | 1.458 (0.019) |
| | Gly | 1.432 | 1.456 | 1.442 (0.008) | 1.451 (0.016) |
| | Pro | 1.462 | 1.467 | 1.465 (0.002) | 1.466 (0.015) |
| *B. Angle* | | | | | |
| $C^{\alpha}$–C–N | Except Gly, Pro | 112.73 | 120.45 | 116.34 (1.53) | 116.2 (2.0) |
| | Gly | 113.73 | 119.79 | 117.63 (1.91) | 116.4 (2.1) |
| | Pro | 115.07 | 116.52 | 115.79 (0.73) | 116.9 (1.5) |
| C–N–$C^{\alpha}$ | Except Pro | 120.38 | 125.42 | 122.87 (1.03) | 123.0 (1.6) |
| | Pro | 122.28 | 123.55 | 122.91 (0.64) | 122.0 (1.4) |
| C–N–$C^{\alpha}$ | Except Gly, Pro | 118.24 | 126.41 | 121.72 (1.60) | 121.7 (1.8) |
| | Gly | 120.09 | 121.94 | 121.06 (0.62) | 120.6 (1.7) |
| | Pro | 119.49 | 119.51 | 119.50 (0.01) | 122.6 (5.0) |
| $C^{\alpha}$–C–O | Except Gly | 116.84 | 123.42 | 120.72 (1.30) | 120.8 (1.7) |
| | Gly | 116.89 | 123.20 | 119.32 (1.74) | 120.8 (2.1) |
| $C^{\beta}$–$C^{\alpha}$–C | Ala | 108.83 | 111.15 | 109.97 (0.94) | 110.5 (1.5) |
| | Ile, Thr, Val | 103.06 | 113.80 | 110.89 (2.04) | 109.1 (2.2) |
| | The rest | 106.43 | 113.84 | 110.24 (1.43) | 110.1 (1.9) |
| N–$C^{\alpha}$–C | Except Gly, Pro | 106.80 | 114.11 | 110.44 (1.89) | 111.2 (2.8) |
| | Gly | 109.62 | 118.19 | 114.60 (2.29) | 112.5 (2.9) |
| | Pro | 110.05 | 110.17 | 110.11 (0.06) | 111.8 (2.5) |
| N–$C^{\alpha}$–$C^{\beta}$ | Ala | 109.74 | 110.75 | 110.21 (0.39) | 110.4 (1.5) |
| | Ile, Thr, Val | 109.01 | 113.50 | 111.14 (1.26) | 111.5 (1.7) |
| | Pro | 102.87 | 103.29 | 103.08 (0.21) | 103.0 (1.1) |
| | The rest | 107.31 | 114.44 | 110.74 (1.48) | 110.5 (1.7) |

[a] The small molecule data used in the above analysis from Ref. 12.
[b] Standard deviation in parentheses.

### Secondary structure and the planarity of the peptide bond

According to an analysis by PROCHECK,[15] 93% of residues other than glycine and proline are in the most favored regions of the Ramachandran plot, while 7% of residues are in additionally allowed regions. In addition, seven glycine and two proline residues are in the favorable regions. Main-chain bond lengths and bond angles within the major conformer are within normal standard deviations (0.011 Å for bond lengths and 1.625° for bond angles). The r.m.s. deviations of main-chain bond length and angle in the alternative N terminus and the β1–β2 loop are 0.016 Å and 2.844°, respectively. Although we kept the restraints for the minor conformer in place throughout the refinement, these values are higher than those associated with the major conformer, reflecting greater uncertainty due to lower occupancy. Mean values for main-chain bond lengths and bond angles in the major conformer, which were refined without any restraints, are comparable with those observed for small molecules[16] (Table 3).

The mean ω angle in the refined model is 178.4°, lower than an ideal value of 180°, or the mean value 179.0° from other protein crystallographic studies,[17,18] but in agreement with several atomic-resolution structure determinations, such as that of a ribonuclease from *Streptomyces aureofaciens*[19] (Figure 2). Eight peptides deviate more from the mean than others. Ser261, located the end of α2 helix, has the lowest ω value (162.2°), corroborated unequivocally by the electron density. Asn230, (ω = 169.4°) is also located at the end of a helix (α1). Two adjacent leucine residues, Leu232 and Leu233, located in the short loop between the α1 helix and β4 strand have ω angles of 189.9° and 165.7°, respectively. Two high ω angles were observed for glycine residues, Gly207 (191.3°) and Gly216 (192.5°) in the turn regions. Another two residues in the minor conformation, Thr198 at the end of N-terminal part and Asp204 in the beginning of β1–β2 loop have high ω values (194.8° and 191.2°). However, due to low occupancy of the minor conformer, these values are not as reliable.

### Side-chain conformers

MacArthur and Thornton[14] found a systematic variation with resolution of the mean values of the χ1 rotamers. We compared the values in our ultra-high-resolution structure to their extrapolation of regression line to ultra-high-resolution. For major conformation, the mean values of *gauche*$^-$, *trans*, and *gauche*$^+$ χ1 rotamers are 67.3° (6.0°), 187.5° (8.6°) and −65.0° (7.3°), respectively. Both *gauche*$^-$ and *gauche*$^+$ value are close to the value expected in high resolution (65.6° and −65.4°), while that of the *trans* rotamer deviates from the expected value of 181.6°. We analyzed χ1 rotamers of serine and leucine residues, which show highly significant correlations of mean value to the resolution. There are seven serine residues in the structure. Three residues have *trans* rotamer and two residues have *gauche*$^-$ rotamer, while other two residues have dual or triple occupancies. This is in good agreement with the observation that there is a higher proportion of *gauche*$^-$ rotamer than that of *gauche*$^+$ rotamer in high-resolution structures. However, the values of *gauche*$^-$ or *gauche*$^+$ do not follow the trend of resolution dependency, although admittedly the sample is too small to draw generalized conclusions. The mean value of three *gauche*$^-$ rotamers is 62.6° and it is lower than the mean value of high-resolution structures (66.3°). One rotamer of Ser259 has higher *gauche*$^+$

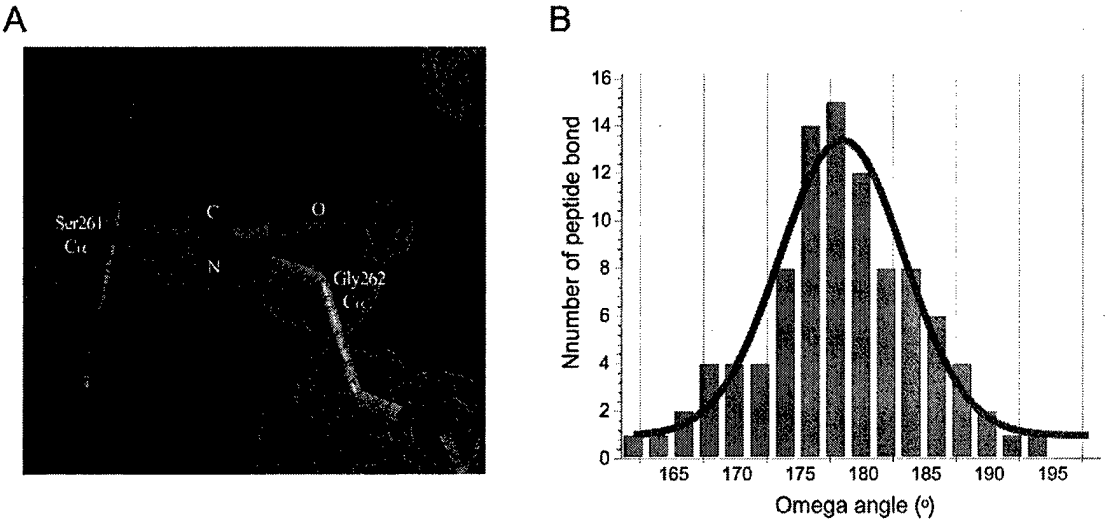



**Figure 2.** The peptide bond distorsions. A, Ser261 with the lowest value (162.2°) of ω angle in the structure: the $2F_o - F_c$ electron density map is contoured at +1.0 σ (gray) and +4.0 σ (red). B, Histogram of ω angles in the refined model.

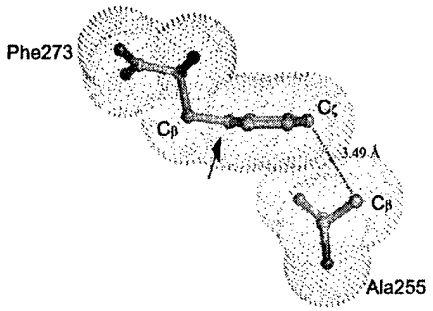



**Figure 3**. Distorted planarity of Phe273. The bending between $C^{\beta}$ and the benzene ring of Phe273 is indicated by an arrow.

value (−54.2°), while the mean value of that in high-resolution structures is −66.8°. Ser205 in major conformation has all three χ1 rotamers as alternative side-chain positions. Values of *gauche*$^-$, *trans* and *gauche*$^+$ rotamers are 60.5°, 176.6° and −57.3°, respectively. Three leucine residues among four in the structure have *gauche*$^+$ rotamers with a mean value of 61.5°. This is in good agreement with the mean value of higher *gauche*$^+$ in high-resolution structures.

There is one case of apparent distorted geometry for a side-chain, i.e. Phe273. The r.m.s. deviation of side-chain atoms from planarity is 0.056 Å, while side-chains of other residues (Arg, Gln, Phe, Asn, Asp, Glu and His) show normal planarity. The phenyl ring of Phe273, the C-terminal residue in this construct, binds in the peptide binding pocket of the adjacent molecule, mimicking a canonical PDZ interaction.[12] The binding pocket includes the glycine-rich carboxylate-binding loop, while the side-chain is embedded within a hydrophobic cluster of Phe211, Phe213, Ala255 and Leu258. The side-chain of Phe273 bends to fit tightly into the pocket, so that $C^{\beta}$ is significantly out of the plane of the phenyl ring (Figure 3). Several atoms of the Phe273 phenyl ring make close contacts ($d < 4.0$ Å) with Leu258, Phe213 and Leu258, but the closest contact is between $C^{\zeta}$ of Phe273 with $C^{\beta}$ of Ala255 (3.49 Å). The latter contact is probably the key stereochemical reason for the bending of the benzene ring. Possible χ1 rotation to relieve this strain is blocked by closely located Ser259 and Met194.

There are eight free carboxyl groups in the structure (aspartate, glutamate and C terminus). As judged by the bond lengths, all are ionized with mean C–O bond of 1.254 Å, ranging from 1.227 Å and to 1.270 Å.

## Hydrogen atoms

Most hydrogen atoms of the model are visible in the electron density map and the inclusion of hydrogen atoms in the model reduces the *R*-factor by 1.6%. We introduced riding hydrogen atoms only into the major conformer. The positions of hydrogen atoms on the rigid group (Y–X–H) are refined with free rotation about the Y–X bond. However, most of the hydrogen atoms in methyl groups lock into unique positions and there is only a limited level of rotation. Hydrogen atoms in CH, NH and aliphatic $CH_2$ groups show good electron density (Figure 4). However, electron density of amide hydrogen in some cases is not clear. Such amide groups are typically involved in the hydrogen bond with an ionized group. The amide groups of Gly210 and Phe211 have hydrogen bonds with the C-terminal carboxyl group of Phe273, and those of Thr206 and His208 interact with the carboxyl group of Asp204. The amide group of Ala193 has a hydrogen bond with the chloride ion in the major conformation of the N terminus. The low electron density of these protons may be due to the strong character of hydrogen bonds and delocalized hydrogen atoms.

## Anisotropic displacement parameters

The anisotropic displacement parameters (*B*-factors) were used for all non-hydrogen atoms, including the water oxygen atoms. The temperature factors of hydrogen atoms were assigned a value 1.2 times greater (1.5 times for methyl group) than that of the heavy atom bound to it. Introduction of the anisotropic model resulted in a significant drop (about 4%) in both $R_{work}$ and $R_{free}$. The mean anisotropy factor for all protein atoms in the major conformer is 0.58, 0.60 for main chain and 0.55 for side-chains. The corresponding average isotropic *B*-value for protein atoms is 6.7 Å$^2$ (4.7 Å$^2$ for main and 7.9 Å$^2$ for side-chain). The mean anisotropy factors for the main chain in flexible parts are lower than average (0.52 for the N terminus and 0.43 for the β1–β2 loop). Many side-chain atoms in minor conformations have large anisotropy (axis ratio is more than 1:5) and some of them are extreme (more than 1:10), probably reflecting genuine dynamic disorder. Our ultra-high-resolution structure has low anisotropy compared to the average anisotropy factor of 0.45 for other anisotropically refined structures.[20]

## Classic hydrogen bonds

There are 187 possible hydrogen bonds, excluding those between water molecules in the structure. These bonds include the interactions where the donor–acceptor distances are less than the radius of acceptor plus 2 Å, and where the D–H–A angle (where D is donor and A is acceptor) is larger than 110°. The mean bond length and angle in this group are 2.94 Å and 154.4°, respectively. With nitrogen as donor, those are 2.95 Å and 155.6°, while with oxygen as donor, those are 2.85 Å and 141.8°. Oxygen is a donor in seven out of 14 cases of short bond length below 2.7 Å compared to 30 out of 187 total hydrogen bonds. Short hydrogen bonds are seen rarely in synPDZ2, and there

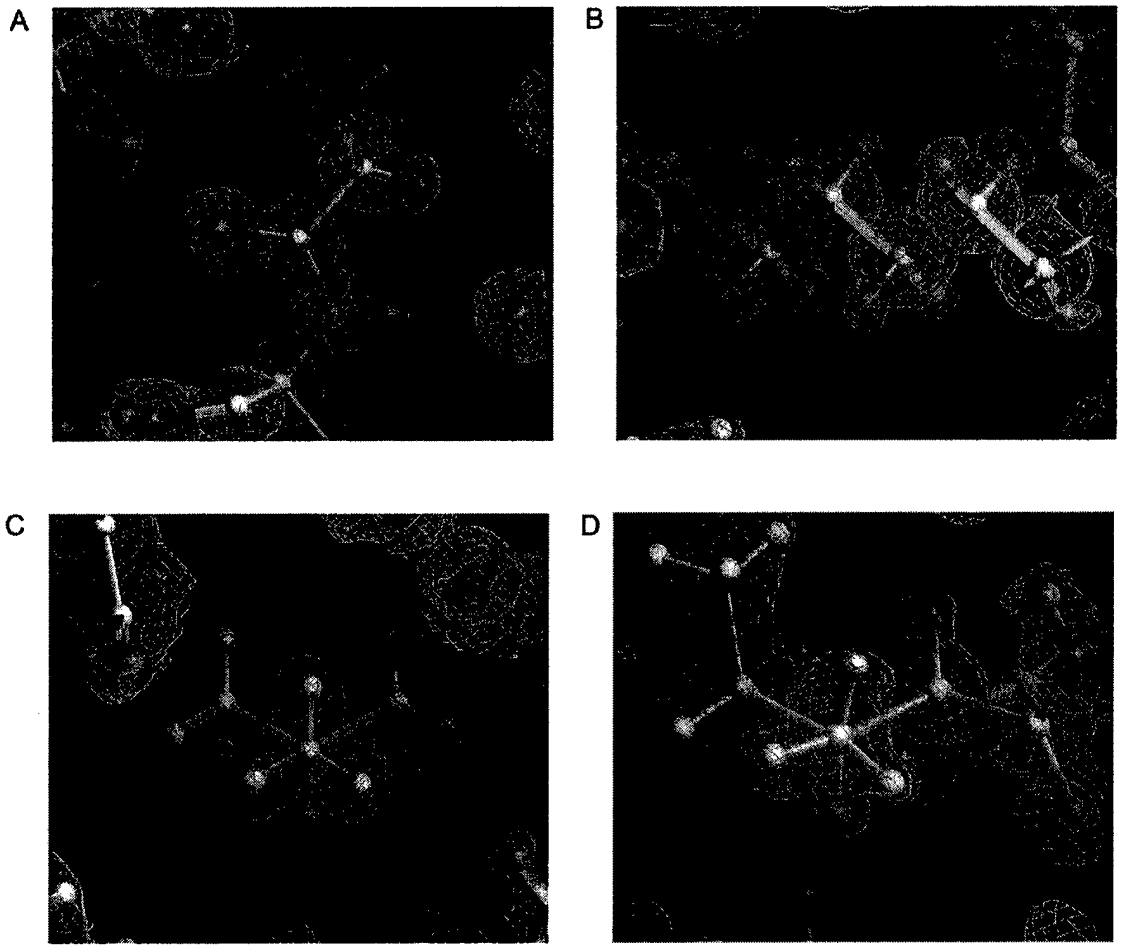



**Figure 4.** Examples of electron density maps revealing the position of hydrogen atoms. The $2F_o - F_c$ electron density map (gray) is contoured at $+1.0\ \sigma$, while the $F_o - F_c$ difference electron density map (red) prior to inclusion of hydrogen atoms is contoured at $+3.0\ \sigma$ and superimposed. A, Ile218; B, Lys203; C, Leu258; and D, Ile247.

are only three shorter than 2.6 Å. None exhibit the characteristics of the so-called low-barrier H-bonds.[10]

### Weak (CH···O) hydrogen bonds

We found several good examples of CH–O hydrogen bonds stabilizing the PDZ-fold. These weak cohesive interactions are seen specifically between β-strands, in accord with previous studies.[21,22] One of them is between Asn237 in the β4 strand and Ile269 in the β6 strand: the carbonyl oxygen atom of Asn237 accepts hydrogen bonds from $C^{\alpha}$ of Ile269 and from the amide group of Met270 (Figure 5). The distance between hydrogen and oxygen in the CH–O bond is closer (2.31 Å) than that in NH bond (2.36 Å), while distances between the nitrogen atom of the amide group to $C^{\alpha}$ and to the carbonyl oxygen atom are virtually identical (3.16 Å and 3.17 Å, respectively). Given the high precision of atomic coordinates in our

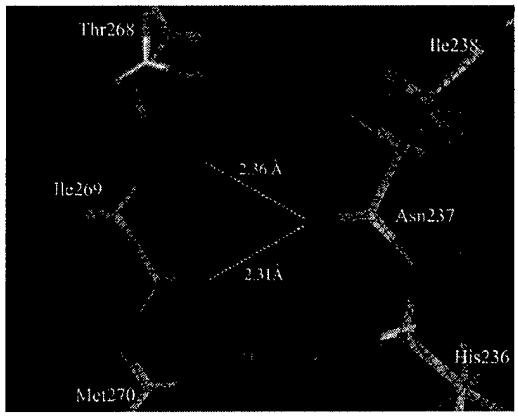



**Figure 5.** The CH–O hydrogen bond between anti-parallel β-strands. The $2F_o - F_c$ electron density map (gray) is contoured at $1\sigma$ and the $F_o - F_c$ difference electron density map (red) of the hydrogen-free model is contoured at $+4.0\ \sigma$ and superimposed.

model, this is a dramatic illustration of the stereochemistry of CH···O bonds in β-sheets.

Another class of weak hydrogen bonds in proteins involves π-acceptors.[23] It includes NH-π or OH-π and weaker CH-π interactions. The π acceptors are typically aromatic rings. However, XH-π hydrogen bonds can be formed with many acceptors other than phenyl rings, such as various heterocycles, C≡C, C=C, and other π bonded moieties.[24] In synPDZ2, we could not find weak hydrogen bonds other than the CH–O bonds, in spite of extensive search using the NCI server.[25]

### Solvent structure

We identified 237 water molecules including partially occupied molecules. A total of 79 water molecules are less than half-occupied and so the total sum of occupancy for water molecules is 150.2. This is equivalent to 300 water molecules per unit cell. Given the solvent content of 31.2% (v/v), about 323 water molecules are possible in a unit cell, and so we identified more than 90% of possible water molecules in the crystal. We also find an additional chloride atom. It is coordinated by the amide groups of Ala193 (3.26 Å) and Ser252 (3.43 Å) of the neighboring molecule $(x, y + 1/2, z)$.

Most water molecules are well ordered with hydrogen bonds to the protein and to other water molecules. The average *B*-factor of water molecules is 10.8 Å$^2$ and, as expected, the distribution of *B*-values depends on the hydrogen bonding pattern. The water molecules with four hydrogen bonds have an average *B* of 5.6 Å$^2$, those with three, two and one hydrogen bonds, show *B*-values of 6.7 Å$^2$, 7.1 Å$^2$ and 8.1 Å$^2$, respectively. The average *B*-value of water molecules beyond the first water shell is 12.8 Å$^2$. Most water molecules with high occupancy have good density for hydrogen atoms involved in hydrogen bonds (Figure 6).

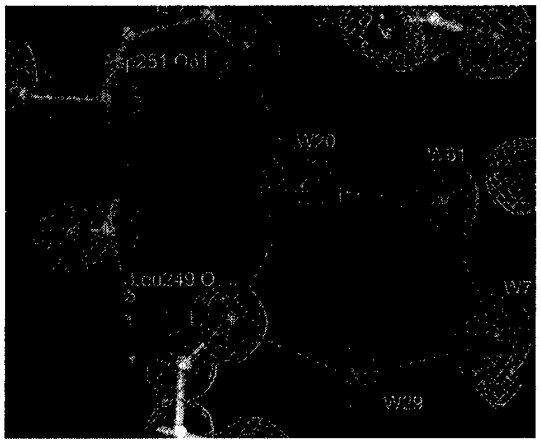



**Figure 6.** Electron density for well-ordered solvent. The $2F_o - F_c$ electron density map (gray) is contoured at 1σ and the $F_o - F_c$ difference electron density map (red) is contoured at +3.5 σ and superimposed.

There are several examples of pentagonal rings of water molecules involving an oxygen atom from the protein.

To compare the positions of water molecules to those from the previous, lower-resolution study (PDB entry 1NTE), we superimposed $C^\alpha$ residues from Thr198 to Met270 of the two models. Among the water molecules within 4 Å of any of the 168 water molecules in the previous structure, 97 are located within 1.2 Å from a water molecule in the low-resolution model, and 51 are within 0.73 Å. Given the precision of the atomic coordinates in both models, this indicates a substantial variation in the solvent structure.

## Discussion

The ultra-high-resolution structure of synPDZ2 narrows the classical discrepancy between small molecule and macromolecular crystallography. For decades, the best protein crystal structures were refined to an *R*-factor in the range 0.15–0.20, while small molecules are routinely refined to 0.02–0.03. This raised the question of what precisely is the cause of the "*R*-gap". The synPDZ2 structure, refined to an *R*-factor of 7.45%, comes near to closing the gap, and it is instructive to reflect why this was possible.

The introduction of an anisotropic vibrational model and inclusion of hydrogen atoms seem to account for no more than the usually observed difference of 0.05–0.07 in the *R*-factor. We note, however, that the inclusion of the minor main-chain conformers in two separate fragments made a critical difference, even though some occupancies are relatively low. When only the main conformer is included, the *R*-factor value is 9.20%. In comparison, elimination of hydrogen atoms from the model increases the *R*-factor to 9.07%, and substitution of isotropic displacement parameters yields an *R*-factor of 12.2%.

The idea of using a limited ensemble of structures for crystallographic refinement is not new, and it was originally proposed by Kuriyan and colleagues, who observed a lower *R*-factor when two structures were used instead of one.[26,27] The result was initially open to the criticism that introduction of more parameters results in over-fitting but, as cross-validation was not yet introduced into crystallographic procedures, the method was never evaluated properly. It might be instructive to revisit this approach, particularly in the case of ultra-high-resolution structures.

The highly refined model of synPDZ2 shows canonical stereochemistry and conforms to the existing libraries of geometric restraints. Two observations are noteworthy: the deviations from planarity of the peptide unit; and the stereochemical evidence for C–H···O hydrogen bonds.

The concept of a planar peptide bond played a key role in the history of protein science. It is generally attributed to Linus Pauling, who may

have had a good understanding of the resonance effects very early on, and who appreciated the importance of the structure of diketopiperazine, which showed a flat peptide unit.[28] He explicitly put the general notion in writing in 1939, in a famous paper co-authored with C. Niemann, demolishing the cyclol theory of protein structure formulated by the British mathematician Dorothy Wrinch.[29] It is less well appreciated that another chemist and an early pioneer of hydrogen bond research, Maurice Huggins, published his thoughts on the resonance effects in the peptide bond in 1937, and showed that this explained the co-planarity of the amide hydrogen with the peptide unit.[30] Interestingly, the concept of the planar peptide unit escaped the attention of many scientists, and as late as in 1950 Bragg *et al.* published a model of a helix in which the peptides were rotated around the C–N bond.[31] Pauling used the planar peptide to predict successfully the α-helix,[32] and ever since the rigid planarity of the peptide bond has featured prominently in all biochemistry and chemistry textbooks. In reality, high-resolution studies show that the bond is relatively flexible allowing for deviations of up to 20° from planarity.[18] This is visible particularly well in the synPDZ2 structure, which shows several significantly twisted peptides with very well resolved electron density. This is of significance for proper application of restraints in crystallographic refinement.

With respect to C–H···O bonds, it is important to realize that chemists have long accepted their existence. As with the peptide bond, the original observation is traced to Pauling, who attributed the high boiling temperature of acetylchloride to a C–H···O bond.[33] The notion that these bonds are ubiquitous in proteins was put forward by Derewenda *et al.*[21,34] and in recent years gained significant support.[35] Stereochemical arguments show that C–H···O bonds may play a particularly significant role in β-sheets, by providing a stabilizing interaction mediated by $C^{\alpha}$ protons, and the free orbitals on the opposing carbonyl oxygen atoms.[22] These interactions, saturating the H-bonding potential of the internal carbonyl group, account for the majority of the so-called lost hydrogen bonds (LHBs) in the core of protein molecules.[36] In spite of the increased acceptance of the notion of weak H-bonds mediated by C–H donor groups, there have been occasional doubts raised if the favorable interactions are not due to artifacts of crystallographic refinement. The synPDZ2 structure, refined with no restraints, provides beautiful examples of stereochemistries consistent with the C–H···O cohesive interactions.

## Materials and Methods

### Crystallization and data collection

The synPDZ2 domain (197–273) from human syntenin was overexpressed as a GST-fusion form and purified as described.[12] Purified synPDZ2 contains an additional pentapeptide (GAMDP) at the N terminus due to the cloning procedure and is designated in the model as residues 192–196. Crystals were obtained in 0.1 M Hepes (pH 7.0) and 34% PEG 4000 using the sitting-drop, vapor-diffusion method with microseeding and were frozen in liquid nitrogen. Data were collected at beamline X8C at NSLS with a wavelength of 0.85 Å using an ADSC Quantum 4R CCD. Two data sets were collected. For high resolution, the detector was set at a distance of 45 mm and the crystal was exposed for 40 seconds with 0.5° of oscillation. For low resolution, the exposure time was ten seconds with 1.0° of oscillation at a detector distance of 90 mm.

Data sets were processed and scaled using HKL2000.[37] The space group is $P2_1$ with unit cell parameters of $a = 25.88$ Å, $b = 39.54$ Å, $c = 32.28$ Å, $\beta = 109.64°$. The completeness of data in the outer shell decreased dramatically due to the shape of the detector. At the resolution range of 20 Å to 0.73 Å, the total number of observations was 316,159 and number of unique reflections was 71,347, yielding an overall completeness of 83.9%.

### Refinement

The structure was solved by the molecular replacement method using AMoRe[38] with starting atomic model (1NTE) and was first refined with a maximum likelihood target function using a subset of data to 1.0 Å resolution with REFMAC5[39] from the CCP4 suite.[40] Manual model building was performed in O.[41] Further refining with all data to 0.73 Å resolution was carried out using SHELX-97[42] with standard conjugate gradient refinement (CGLS). Crystallographic details including data and refinement statistics are shown in Table 2. Structural analysis was carried out with the program PROCHECK,[15] EdPDB[43] and PARVATI[20] and NCI.[25]

### Protein Data Bank accession number

The coordinates and the structural factors have been deposited with the RCSB Protein Data Bank under accession number 1R6J.

## Acknowledgements

Data collection for this study at the NSLS (Brookhaven National Laboratory, Upton, New York) was carried out during RapiData2003, a practical course in macromolecular X-ray diffraction measurement. Supported by the DOD grant DAMD17-01-1-0720.

## References

1. Bernal, J. D. & Crowfoot, D. (1934). X-ray photographs of crystallin pepsin. *Nature*, **133**, 794–795.
2. Phillips, J. C., Wlodawer, A., Yevitz, M. M. & Hodgson, K. O. (1976). Applications of synchrotron radiation to protein crystallography: preliminary results. *Proc. Natl Acad. Sci. USA*, **73**, 128–132.
3. Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R.

(1984). The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* **175**, 159–174.
4. Hough, E., Hansen, L. K., Birknes, B., Jynge, K., Hansen, S., Hordvik, A. *et al.* (1989). High-resolution (1.5 Å) crystal structure of phospholipase C from *Bacillus cereus*. *Nature*, **338**, 357–360.
5. Hope, H. (1988). Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallog. sect. B*, **44**, 22–26.
6. Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1995). Proteins at atomic resolution. *Curr. Opin. Struct. Biol.* **5**, 784–790.
7. Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). The benefits of atomic resolution. *Curr. Opin. Struct. Biol.* **7**, 681–688.
8. Schmidt, A., Gonzalez, A., Morris, R. J., Costabel, M., Alzari, P. M. & Lamzin, V. S. (2002). Advantages of high-resolution phasing: MAD to atomic resolution. *Acta Crystallog. sect. D*, **58**, 1433–1441.
9. Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
10. Cleland, W. W., Frey, P. A. & Gerlt, J. A. (1998). The low barrier hydrogen bond in enzymatic catalysis. *J. Biol. Chem.* **273**, 25529–25532.
11. Hung, A. Y. & Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. *J. Biol. Chem.* **277**, 5699–5702.
12. Kang, B. S., Cooper, D. R. Y. D., Derewenda, U. & Derewenda, Z. S. (2003). Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. *Structure*, **11**, 845–853.
13. Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein? Nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871.
14. MacArthur, M. W. & Thornton, J. M. (1999). Protein side-chain conformation: a systematic variation of chi 1 mean values with resolution—a consequence of multiple rotameric states? *Acta Crystallog. sect. D*, **55**, 994–1004.
15. Laskowski, R., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
16. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.
17. EU 3-D Validation Network (1998). Who checks the checkers? Four validation tools applied to eight atomic resolution structures. *J. Mol. Biol.* **276**, 417–436.
18. MacArthur, M. W. & Thornton, J. M. (1996). Deviations from planarity of the peptide bond in peptides and proteins. *J. Mol. Biol.* **264**, 1180–1195.
19. Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). Ribonuclease from Streptomyces aureofaciens at atomic resolution. *Acta Crystallog. sect. D*, **52**, 327–344.
20. Merritt, E. A. (1999). Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta Crystallog. sect. D*, **55**, 1109–1117.
21. Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). The occurrence of C–H···O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248–262.
22. Lee, K. M., Chang, H. C., Jiang, J. C., Chen, J. C., Kao, H. E., Lin, S. H. & Lin, I. J. (2003). C–H···O hydrogen bonds in beta-sheetlike networks: combined X-ray crystallography and high-pressure infrared study. *J. Am. Chem. Soc.* **125**, 12358–12364.
23. Steiner, T. & Koellner, G. (2001). Hydrogen bonds with pi-acceptors in proteins: frequencies and role in stabilizing local 3D structures. *J. Mol. Biol.* **305**, 535–557.
24. Brandl, M., Weiss, M. S., Jabs, A., Suhnel, J. & Hilgenfeld, R. (2001). C–H···pi-interactions in proteins. *J. Mol. Biol.* **307**, 357–377.
25. Babu, M. M. (2003). NCI: a server to identify non-canonical interactions in protein structures. *Nucl. Acids Res.* **31**, 3345–3348.
26. Kuriyan, J., Brunger, A. T., Karplus, M. & Hendrickson, W. A. (1989). X-ray refinement of protein structures by simulated annealing: test of the method on myohemerythrin. *Acta Crystallog. sect. A*, **45**, 396–409.
27. Kuriyan, J., Osapay, K., Burley, S. K., Brunger, A. T., Hendrickson, W. A. & Karplus, M. (1991). Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins: Struct. Funct. Genet.* **10**, 340–358.
28. Corey, R. B. (1938). The crystal structure of diketopiperazine. *J. Am. Chem. Soc.* **60**, 1598–1604.
29. Pauling, L. & Niemann, C. (1939). The structure of proteins. *J. Am. Chem. Soc.* **61**, 1860–1867.
30. Huggins, M. L. (1937). Hydrogen bridges in organic compounds. *J. Org. Chem.* **1**, 407–456.
31. Bragg, L., Kendrew, J. C. & Perutz, M. F. (1950). Polypeptide chain configurations in crystalline proteins. *Proc. Roy. Soc. ser. A*, **203**, 321–357.
32. Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the peptide chain. *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
33. Pauling, L. (1960). *The Nature of the Chemical Bond*, 3rd edit., Cornell University Press, Ithaca, NY.
34. Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). (His)C ε–H···O=C< hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.* **241**, 83–93.
35. Wahl, M. C. & Sundaralingam, M. (1997). C–H···O hydrogen bonding in biology. *Trends Biochem. Sci.* **22**, 97–102.
36. Savage, H., Elliot, C., Freeman, C. & Finney, J. (1993). Lost hydrogen bonds and buried surface area: rationalizing stability in globular proteins. *J. Chem. Soc. Faraday Trans.* **89**, 2609–2617.
37. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
38. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
39. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
40. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
41. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of

errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
42. Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: higher resolution refinement. *Methods Enzymol.* **227**, 319–343.
43. Zhang, X. & Matthews, B. W. (1995). EDPDB: a multifunctional tool for protein structure analysis. *J. Appl. Crystallog.* **28**, 624–630.

***Edited by R. Huber***

*(Received 29 October 2003; received in revised form 14 February 2004; accepted 20 February 2004)*

Structure, submitted Oct 5, 2004

# Probing the supramodular architecture of a multidomain protein: the solution structure of syntenin

Tomasz Cierpicki[1], John H. Bushweller[1,2] and Zygmunt S. Derewenda[1]

[1]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908

[2]Department of Chemistry, University of Virginia, Charlottesville, Virginia 22906

**Corresponding Author**

Zygmunt S. Derewenda

Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908; Phone: (434) 982-3151; FAX: (434) 982-1616;

e-mail:zsd4n@virginia.edu

**Running Title**

Solution structure of syntenin.

**Keywords**

Syntenin, domain orientation, residual dipolar couplings, NMR

## Summary

Full understanding of the mechanisms of function of multidomain proteins requires knowledge of their supramodular architecture in solution. This is a non-trivial task for either X-ray crystallography and NMR, because intrinsic flexibility makes crystallization of these proteins difficult, while their size creates a challenge for NMR impractical. We here describe synergistic application of data derived from X-ray crystallography and residual dipolar couplings (RDCs), to address the question of the supramodular structure of syntenin, a 32 kDa protein containing two PDZ domains and involved in cytoskeleton-membrane organization. We show that the mutual disposition of the PDZ domains clearly differs from that seen in the crystal structure and we provide evidence that N- and C-terminal fragments of syntenin, hitherto presumed to lack ordered structure, contain folded structural elements in the full-length protein in contact with the PDZ tandem.

## Introduction

A large fraction of genes in eukaryotic genomes codes for large, multidomain proteins, many of which are critically responsible for complex pathways of cell regulation. While X-ray crystallography and NMR have been successful at structural characterization of isolated domains and their binary complexes, the task of characterization of solution structures of full-length, multidomain proteins, is an immense challenge to the two techniques. The intrinsic flexibility of most multidomain proteins makes it very difficult

to grow diffracting single crystals, unless they can be induced to form a 'closed', compact conformation. There is also the danger that crystal packing forces may distort the structure of a protein with a high degree of intrinsic flexibility, or stabilize a conformer which is not representative of the structure in solution. On the other hand, while not dependent on crystalline samples, NMR methodology becomes very labor-intensive and expensive with increasing molecular weight of the protein. Clearly, a better understanding of the behavior of multidomain proteins in solution requires development of alternative approaches, including a synergistic use of X-ray diffraction models and NMR-derived data. Such an approach has been made possible with the introduction of residual dipolar couplings (RDCs) as an alternative and complementary approach greatly expanding the scope of NMR methodology (Prestegard, et al., 2000, Bax, et al., 2001, de Alba and Tjandra, 2002, Bax, 2003). The RDCs make it possible to determine the orientation of specific bonds, such as $^{1}H$-$^{15}N$ peptide moieties for partially oriented molecules. While the determination of RDCs still requires full assignment of chemical shifts, the spectra can be analyzed in a significantly shorter time than standard NOE based experiments. Furthermore, unlike the chemical shifts which cannot be predicted with very high accuracy, RDCs can easily be computed based on a set of crystallographic coordinates, making it easy to combine crystallographic and NMR data to generate a comprehensive description of structural properties of a given protein in solution.

In this report we describe the use of this novel approach to characterize the solution supramodular architecture of syntenin, a 32 kDa scaffolding protein which contains two PDZ domains arranged in a tandem, and N- and C-terminal fragments, hitherto presumed to be unstructured based on secondary structure prediction. PDZ

domains—between 80 and 90 amino acids in size—are the most ubiquitous protein-protein interaction domains, and are found exclusively in cytoplasmic, multidomain proteins often in conjunction with one or more domains of other types (Nourry, et al., 2003). They typically function by selective binding of C-terminal oligopeptides of other proteins, notably channels and receptors (Doyle, et al., 1996, Hillier, et al., 1999), and thus play a critical role in the organization of signaling complexes (Fan and Zhang, 2002). Syntenin was originally identified as a binding partner of the cytoplasmic domain of vertebrate syndecans (Grootjans, et al., 1997). It is widely expressed, strongly associated with membranes, and plays a role in cytoskeleton-membrane organization (Zimmermann, et al., 2001), protein trafficking, cell adhesion, and activation of the transcription factor Sox4 (Geijsen, et al., 2001), through interactions with a number of other proteins including ephrins (Lin, et al., 1999) and neurexins (Lin, et al., 1999, Grootjans, et al., 2000). The PDZ tandem of syntenin and the isolated second (PDZ2) domain have been studied by X-ray crystallography (Kang, et al., 2003, Kang, et al., 2003, Kang, et al., 2004). These studies revealed that the PDZ tandem forms a head-to-tail homodimer, suggesting that dimerization may be a biologically relevant phenomenon, in concert with some other biochemical data (Koroll, et al., 2001). The crystal structure showed a rigid supramodular architecture, with a possibility of domain-swapping within the homodimer (Figure 1).

Since the relative disposition of PDZ domains within multi-PDZ proteins is increasingly recognized as having an important biological role, our study was intended to answer the following unresolved issues: (a) is full length syntenin dimeric or monomeric in solution? (b) can we detect domain swapping events in solution? (c) do the domains

retain in solution the supramodular architecture observed in the crystal? (d) do the N- and C-termini, presumed to be unstructured, contain any folded elements interacting with the PDZ tandem?

By combining the NMR derived information, based on the RDC measurements, with the known crystal structures, we show that in solution syntenin is monomeric but its supramodular structure is significantly different from that seen in the crystal. We also present evidence that the C-terminus affects the structure of the tandem while the N-terminal fragment is unstructured. This investigation provides a powerful illustration of the complementary role of crystallography and solution NMR, and shows how the combination of these techniques allows one to overcome the limitations of each of these methods when applied in isolation.

## Results

### Chemical shift analysis of syntenin PDZ domains

As the first step in the study of the supramodular structure of syntenin in solution, we needed to compare chemical shifts of the two isolated PDZ domains and that of the tandem. Each of the two PDZ domains, PDZ1 and PDZ2, were individually expressed and purified (see Materials and Methods). As a prerequisite for the RDC analysis, it was necessary to obtain reasonably complete chemical shift assignments for amides, using samples of $^{15}N$ labeled proteins. This was accomplished based on the analysis of $^{15}N$-edited HSQC-TOCSY, $^{15}N$-edited HSQC-NOESY and HNHA experiments. The NMR $^{1}H$-$^{15}N$ HSQC spectra of both isolated PDZ1 and PDZ2 domains are well dispersed.

Assignments of $^{1}H^{N}$, $^{15}N$ and $^{1}H^{\alpha}$ chemical shifts were 100 % and 90 % complete for PDZ1 and PDZ2, respectively. Interestingly, the chemical shift assignment of the PDZ2 domain revealed that the backbone amide of Asn215 is strongly downfield-shifted ($\delta_{HN}$=12.66 ppm). Since there are no aromatic rings in the proximity of this amide, the unusually large chemical shift (one of the largest amide chemical shifts reported in the literature) must result from a very short hydrogen bond with the side chain of Asp251. This is further corroborated by a large temperature coefficient for the amide of Asn215 ($\Delta\delta_{HN}/\Delta T$= -5.8ppb/K) (Cierpicki, et al., 2002). The crystal structure (PDB code 1n99) shows that the distance between the backbone nitrogen of Asn215 and the side chain oxygen of Asp251 is 2.56Å.

In order to compare the isolated PDZ domains to the PDZ tandem in solution, we also measured the $^{1}H$-$^{15}N$ HSQC spectra for the tandem. The comparisons of the spectra indicate that most of the amide chemical shifts do not vary significantly between the isolated domains and the tandem. Thus, the knowledge of the chemical shifts for the isolated domains was of significant help in the assignment of the PDZ tandem, and consequently 96 % of the $^{1}H^{N}$, $^{15}N$ and $^{1}H^{\alpha}$ resonances, including those corresponding to the linker region between the PDZ domains, were assigned.

The observed differences in $^{1}H^{N}$ and $^{15}N$ chemical shifts between isolated domains and the tandem are summarized in Figure 1. Interestingly, much stronger changes are observed for PDZ2 relative to PDZ1 and the largest differences exist for PDZ2 residues 213-218 and 234-241. The crystal structure of the PDZ tandem shows that three aromatic rings (Phe154, Phe195 and Phe273) are buried in the surface between the PDZ domains and thus ring current effects may be the source of some of the observed

chemical shift changes. However, in spite of these localized differences, there are no substantial changes for residues within close proximity to the peptide binding sites.

**The PDZ tandem in solution**

The crystal structure of PDZ tandem is consistent with a homodimeric structure, arranged head-to-tail, so that PDZ1 in one monomer makes extensive contacts with PDZ2 of the other monomer in the asymmetric unit (Kang, et al., 2003). Probing whether the dimeric or monomeric state is predominant in solution we carried out measurement of $^{15}N$ relaxation times. The estimated correlation times, inversely proportional to the tumbling rate of the molecules in solution, determined for PDZ1, PDZ2 and for the tandem from $^{15}N$ relaxation time ratios (T1/T2) are 4.7±0.2, 4.5±0.5 and 10.0±0.5 ns, respectively. The predicted values, assuming an axially symmetric model for the tandem are 13 and 26 ns for a monomeric and dimeric species, respectively (see Materials and Methods). Thus, the experimental NMR data for the tandem are consistent with a predominantly monomer state of the protein in solution.

In order to gain qualitative insight into the backbone mobility within the tandem, we compared the $^{15}N\{^{1}H\}$ NOEs for each isolated domain with those measured for the tandem. A relatively uniform distribution of NOEs was observed with average values of 0.77±0.05, 0.76±0.03 and 0.75±0.08 for PDZ1, PDZ2 and PDZ12, respectively. Interestingly, the NOE values for the residues within the linker region between PDZ1 and PDZ2 (i.e. Phe195, Glu196 and Arg197) are not diminished. This strongly suggested that the two PDZ domains have fixed relative orientations within the tandem.

**The relative orientation of the PDZ domains in the tandem**

Next, we determined the mutual dispositions of the two PDZ domains in the tandem, based on the RDCs. To measure the latter, several samples with different gel compositions were used to find optimal conditions for inducing the weak alignment. We observed an alignment of the tandem in compressed copolymer polyacrylamide gels (see Materials and Methods). Two sets of $^1D_{HN}$ values were measured, and the experimental values of RDCs ranged from –14 to +15 Hz and –11.5 to +14.5 Hz for 50+M and 75+M, respectively. However, the correlation coefficient between the two data sets (r) is 0.97 and the alignments are virtually identical. Qualitative compatibility between dipolar couplings and the corresponding crystal structures were evaluated using the quality factor Q (Cornilescu, et al., 1998). A value of zero indicates a perfect fit, while in reality a value of approximately 20% correspond to a very good fit between the two sets of observations (see Materials and Methods) (Ottiger and Bax, 1999). In order to select structures of the component PDZ domains that most accurately represent the structure in solution we compared the quality factors (Q). Values of Q calculated for individual domains from PDZ12 using the 75+M data set are 22.3, 22.1, 28.5 and 22.5 % for $PDZ1^A$ $PDZ2^A$, $PDZ1^B$ and $PDZ2^B$, respectively (superscripts denote monomers A and B from the tandem crystal structure). Thus, it is most likely that monomer A constitutes the best representation of the structure in solution. Parameters of the alignment tensors were obtained from the best fit of measured and calculated RDCs based on the crystal structures of $PDZ1^A$ and $PDZ2^A$ domains (see Table I). Interestingly, the alignment of the individual domains is very similar, although the orientations of PDZ1 and PDZ2 are noticeably different. The quality factor calculated for the crystal structure of tandem

(PDZ12$^{A}$) is 27.7 %, significantly higher than that of the individual domains, indicating a possibility of altered interdomain orientation in solution.

We then used the RDCs for each individual PDZ domain to determine the relative orientations of these modules in the PDZ tandem in solution. We first aligned the coordinates of the PDZ tandem to the alignment tensor frame of PDZ1, rotating the crystal structure by –26°, 18° and –42° about x, y and z axes, respectively. Under those circumstances, the alignment tensor of PDZ2 is rotated with respect to that of PDZ1 by -5°, 3° and -23° around the x, y and z axes respectively (Table I). Application of the above transformation to PDZ2 yielded the architecture of PDZ tandem that is consistent with the dipolar couplings recorded in solution. The resulting quality factor calculated for the tandem upon reorientation of the PDZ domains dropped from 27.7 to 22.7 % and is similar to that calculated for the individual domains. The magnitude of the angular errors in domain orientation calculated from the residual dipolar couplings has been evaluated using the jack-knife procedure (Mosteller and Tukey, 1977) and through a comparison of two independent sets of dipolar couplings. The two data sets are very similar and the differences do not exceed 1.5° for rotation about each of the angles.

Since the analysis of the RDCs from a single alignment gives rise to a fourfold degeneracy in the fragment orientations, we had to consider the possibility of an alternate interdomain arrangement. Thus, we constructed domain swapped tandem represented by PDZ1$^{A}$ and PDZ2$^{B}$ and we assessed the compatibility of this model with the experimentally determined RDCs. The agreement was poor, as judged by the high value of the quality factor (72.8 %). Furthermore, the domain swapped structure is less consistent with the chemical shift data (Figure 1).

**Supramodular structure of PDZ tandem**

Although the RDCs allow for an accurate determination of the relative orientations of the domains, they do not provide information regarding the relative translations. To obtain a complete structural description of the tandem it is necessary to obtain additional information with respect to the interdomain interface. Such information can be derived from the comparison of chemical shifts between isolated PDZ domains and the PDZ tandem (Figure 1) and their use in the form of ambiguous distance restraints for structure calculations (Clore and Schwieters, 2003).

Two independent simulations employing two sets of dipolar couplings resulted in very similar structures of the tandem (Figure 2A). An extensive set of intradomain restraints allowed us to preserve a rigid backbone structure during the simulation. The root-mean-square deviations between main chain atoms of the ten lowest energy conformers of PDZ1 and PDZ2 relatively to crystal structure are 0.10 Å and 0.17 Å, respectively. While the relative orientation of PDZ domains was restricted during the simulation due to the use of dipolar couplings, the key point to obtain a compact structure of the PDZ12 was the application of ambiguous distance restraints derived from chemical shifts analysis. The ambiguous distance restraint was not violated more than 0.2 Å in the final structures of PDZ12 and there are no steric clashes at the interface between the PDZ domains. Nevertheless, we observed that in three out of ten structures calculated using the 75+M set of RDCs, one domain is translationally shifted relative to the second domain by 4 Å (not shown). Highly ambiguous distance restraints constructed on the basis of chemical shift comparison do not discriminate between interdomain contact and

small intradomain structural changes. Thus, they may contain erroneous information eventually leading to a set of incorrect structures. Such errors can be eliminated by inspecting all sets of calculated structures. The translational movement is very unlikely to occur since it decreases the contact area between the PDZ domains. The three incorrect structures were accordingly eliminated from further analysis. Such structures have not been observed in calculations using a second set of RDC (50+M). Comparison of the tandem structures calculated using 50+M RDC data (magenta) and 75+M (green) is shown in Figure 2A.

A superposition of a representative solution conformer of PDZ12 (blue) and the crystal structure of PDZ12_A (green - monomer A from the crystal structure) is shown in Figure 2B. Relative rotation of PDZ2 is clearly visible and root-mean-square deviation between main chain atoms of the two PDZ2 domains is 3.9 Å. As a consequence a number of backbone atoms are translationally shifted by more than 5 Å relative to the crystal structure (calculate RMSD).

In order to examine whether isolated PDZ domains are able to interact in solution we performed an NMR titration experiment in which unlabelled PDZ2 was added to $^{15}N$ labeled PDZ1 in a 1:1 molar ratio (data not shown). Since no chemical shifts of PDZ1 were affected, we concluded that the covalent linker between the domains is necessary to maintain the contact between the PDZ domains.

**The N- and C-termini**

Structural analysis of modular proteins is often limited to domains with well defined tertiary structures. While the protein-protein interactions are frequently mediated by these

globular domains, the presence of full-length intact proteins including unfolded fragments is necessary for the expression of full biological function. Detailed study of such unstructured fragments is difficult. In the case of syntenin over 45 % of the polypeptide chain extends beyond the domain boundaries. In order to address the question of whether the terminal fragments of the protein form any contacts with PDZ1 and PDZ2 domains, possibly modulating their structure and interactions, we prepared $^{15}$N-labeled full-length syntenin (FL). Despite the large size of the FL protein, the $^{1}$H-$^{15}$N HSQC spectrum shows relatively narrow signals corresponding to the PDZ12 fragment. Additionally, the severely overlapped region in the center of the spectrum is characteristic for the unstructured N- and C-termini (Figure 3). Interestingly, the detailed comparison of full length protein and PDZ12 spectra indicates distinct differences including an increase in the number of signals in the folded region (see below).

In order to evaluate the effects of the N- and C-termini, we analyzed two additional syntenin constructs: N-PDZ12 (residues 1-273) and PDZ12-C (residues 113-298). Comparison of $^{1}$H-$^{15}$N HSQC spectra for all syntenin fragments measured under identical conditions is shown in Figure 3. Spectrum of N-PDZ12 is very similar to the PDZ tandem and contains a number of additional resonances arising from the N-terminus (as seen for FL in Figure 4). All these extra signals are found to lie within random-coil positions indicating that this fragment is unstructured. However, addition of the N-terminal extension causes numerous shifts for resonances within the PDZ12 fragment (Figure 4A). In order to unambiguously assign these changes we prepared an additional sample by mixing equimolar amounts of $^{15}$N-labelled N-PDZ12 and PDZ12. In such an experiment, two peaks with diminished intensities should be observed for all amides that

differ in chemical shifts between PDZ12 and N-PDZ12. We analyzed all signals that were not overlapped with the N-terminus and found that all amides with affected chemical shifts belong to PDZ1 and lie in close proximity to the N-terminus. Thus, in spite of the fact that numerous amides have been affected it is most likely that all perturbations of chemical shifts are caused by the attachment of the N-terminal polypeptide that is most likely unstructured and does not interfere with the PDZ tandem.

Unlike the N-PDZ12, the spectrum of PDZ12-C shows numerous strongly broadened resonances (Figure 4B). Furthermore, the sample of PDZ12-C is less stable in solution and tends to precipitate even at low concentration. The observed signal broadening may indicate an increased aggregation propensity of the protein upon addition of the 25 C-terminal residues to the tandem. The spectrum of PDZ12-C indicates numerous chemical shift changes relative to the PDZ tandem alone (Figure 4B); however, due to strong signal broadening a reliable assignment of the affected residues is not possible. Interestingly, full length syntenin is more stable relative to PDZ12-C and a high resolution spectrum was measured (Figure 3). A comparison of the $^{1}H$-$^{15}N$ HSQC spectra of the full length protein and PDZ12 reveals broader signals (Figure 4C), but this effect can be attributed to a larger molecular weight of full length syntenin (32 kDa relatively to 19 kDa for PDZ12). The most interesting feature of FL spectrum is that roughly all resonances are shifted relative to PDZ12. Furthermore, we could identify additional signals indicating that most likely several residues outside the PDZ domains are structured as well. For example, the spectrum of the full-length protein reveals the presence of an upfield-shifted amide with a proton chemical shift of 6.34 ppm (Figure 3). The comparison of the different fragments and the full length protein shows that the

presence of both termini is necessary for maintaining the proper structure of syntenin. It is likely that residues from the C-terminal fragment of syntenin as well as possibly the N-terminus are structured and interact with the PDZ domains.

**Discussion**

Regulatory and scaffolding proteins often contain multiple PDZ domains, frequently closely spaced along the sequence (Jelen, et al., 2003, Nourry, et al., 2003). Such architecture is important for their function, as has been demonstrated—for example—for the first two PDZ domains of PSD-95, a protein mediating ion channel clustering (Imamura, et al., 2002). At least some multi-PDZ proteins have defined supramodular architectures and their interactions with binding partners display a higher level of complexity (Feng, et al., 2003, Im, et al., 2003, Long, et al., 2003). Furthermore, sequence analyses indicate that PDZ domains are frequently connected by short linkers of conserved length, which may impose constraints on the mutual disposition of the adjacent domains. The structure of the PSD-95 PDZ domain tandem revealed restricted orientation of PDZ domains leading to concerted orientation of peptide binding sites (Long, et al., 2003). Similarly, the solution structure of the tandem of PDZ4 and PDZ5 from the glutamate receptor interacting protein (GRIP) revealed that PDZ domains form a compact structure with a fixed interdomain orientation (Feng, et al., 2003). Supramodular architecture may also result from multimerization of PDZ domains. The recently reported structure of the Shank PDZ domain shows a tightly associated dimer that is maintained both in solution and in the crystal (Im, et al., 2003). A dimeric association may facilitate binding to dimeric ligands such as βPIX (Im, et al., 2003). Despite their small size and

simple structure, PDZ domains display a broad spectrum of interaction modes, the knowledge of which is crucial for the understanding of the architecture of signaling complexes.

In this report, we address the question of the domain orientation and supramodular structure of the syntenin's PDZ tandem, as determined using residual dipolar couplings and chemical shifts. Our study shows that the PDZ tandem is monomeric in solution and that the two domains tumble as a single unit with a rotational correlation time of 10 ns. The mutual arrangement of the PDZ domains in solution has also been determined from residual dipolar couplings. While it is similar to that seen in the crystal structure, the domains are rotated approximately -5°, 3°, -23° about the x, y, z axes, respectively (Figure 2B). This difference leads to a rearrangement of the buried surfaces at the PDZ1/PDZ2 interface. In the crystal structure of PDZ tandem this interface is actually much smaller than the intermolecular contact, i.e. the interface between the PDZ1 and PDZ2 that belong to the different monomers (Kang, et al., 2003). While, the intramolecular surface area buried between the two PDZ domains within monomers A and B is 283 and 340 $Å^2$, respectively, the intermolecular contacts are almost twice as extensive (547 $Å^2$ and 573 $Å^2$ for $PDZ1^A$-$PDZ2^B$ and $PDZ1^B$-$PDZ2^A$, respectively). Thus, reorientation of the domains observed for the solution structure is most likely driven by a significant increase of the intramolecular contact surface between PDZ1 and PDZ2, which we estimate using two independent sets of dipolar couplings to be ~480 $A^2$. Clearly, the crystallization process and crystal packing forces select a conformation which is significantly different from the one in solution. Similar effects have been reported in other cases. For example, the solution structure of lysozyme derived from RDCs differs from the crystal structures

and the cleft between the two domains is significantly larger in solution (Goto, et al., 2001). Similarly, the solution state of MBP/β-cyclodextrin complex has been found to be ~10° more closed than the crystalline state (Evenas, et al., 2001). In it clear that synergistic application of residual dipolar couplings and crystal structures of individual domains leads to a much more accurate determination of the architecture of modular proteins in solution.

Our work also illustrates the structural consequences of the presence of partly unfolded fragments in the protein. This is supported by the NMR spectra of the full-length syntenin. While the addition of the N-terminal fragment affects only signals within PDZ1, the presence of the C-terminal fragment leads to a strong chemical shift changes accompanied with a severe signal broadening effect as a result of aggregation. The full length syntenin has a noticeably lower tendency to aggregate and consequently high resolution spectra can be recorded. The comparison of the $^{1}H$-$^{15}N$ HSQC spectra of the full-length protein relative to those of the tandem reveals an increased number of amide resonances with non-random coil chemical shifts. This indicates that there are fragments extraneous to the PDZ domains that are structured in the full-length protein. It is likely that this involves the C-terminal fragment, although an interaction with the N-terminus should not be ruled out. This observation is in agreement with thermodynamic studies previously reported for syntenin which show that the full length protein unfolds cooperatively and is more stable by 2 kcal/mol relative compared to the isolated tandem (Kang, et al., 2003). Whether or not this effect modulates syntenin's function and/or peptide binding requires further study.

## Experimental Procedures

### Theory

The residual dipolar couplings (RDCs) emerge from incomplete averaging of dipole-dipole interactions in solution NMR and provide information about orientation of internuclear vectors relative to the magnetic field (Prestegard, et al., 2000, Bax, et al., 2001, de Alba and Tjandra, 2002, Bax, 2003). RDC between pairs of nuclei is defined by:

$$D_{IS} = \frac{-\mu_0 h \gamma_I \gamma_S}{16\pi^3 r_{IS}^3}\left[A_a\left(3\cos^2\theta - 1\right) + \frac{3}{2}A_r \sin^2\theta\cos 2\phi\right]$$

where $\gamma_I$ and $\gamma_S$ are gyromagnetic ratios of interacting nuclei, $r_{IS}$ is a distance between I and S, $\theta$ and $\phi$ describe orientation of IS intermolecular vector relatively to molecular alignment tensor and $A_a$ and $A_r$ are axial and rhombic components of the alignment tensor, respectively. The dipolar coupling between a pair of $^1H$ and $^{15}N$ nuclei can be simplified as follows:

$$D_{IS} = D_a\left[\left(3\cos^2\theta - 1\right) + \frac{3}{2}R\sin^2\theta\cos 2\phi\right]$$

where $D_a$ is the magnitude of the alignment tensor normalized to interaction between $^1H^N$ and $^{15}N$ nuclei and R is rhombicity. Parameters of the alignment tensor ($D_a$ and R) can be obtained from best fitting of experimental RDC to the known structure (Losonczi, et al., 1999). This procedure additionally yields the Euler angles ($\alpha$, $\beta$, $\gamma$) defining rotation of molecular coordinates relative to the alignment tensor frame. Thus, a combination of RDCs and structures determined by either NMR or crystallographic analysis may be conveniently used to establish the relative orientation of molecular fragments in solution (Fischer, et al., 1999).

Measurement of RDCs is possible only under the conditions of anisotropic tumbling of molecules in solution. So far, the most successful methods used for obtaining weak alignment of macromolecules are based on diluted liquid crystalline media (Bax and Tjandra, 1997) and filamentous phages (Clore, et al., 1998). More recently, strained polyacrylamide based hydrogels were introduced as inert and very stable orienting media (Sass, et al., 2000, Ishii, et al., 2001). Further modifications of the gel composition makes possible the preparation of charged copolymers that are stable at low concentrations and are suitable for studying large proteins (Meier, et al., 2002, Ulmer, et al., 2003).

The orienting media should be carefully chosen in order to avoid strong interactions with the protein. In our initial studies we were unsuccessful with the DMPC:DHPC bicelles (Bax and Tjandra, 1997) or negatively charged polyacrylamide copolymer gels (Meier, et al., 2002), due to very strong broadening of protein signals. However, we were able to utilize the recently developed positively charged copolymer gels with composition of 50% (3-acrylamidopropyl)-trimethylammonium chloride - 50% acrylamide (referred to hereafter as 50+M) and 75% (3-acrylamidopropyl)-trimethylammonium chloride - 25% acrylamide (75+M).

**Protein purification**

Six different syntenin fragments have been prepared: full length protein (1-298), N-PDZ12 (1-273), PDZ1 (113-193), PDZ2 (197-273), PDZ12-C (113-298). All proteins were uniformly $^{15}N$-labeled by overexpression in minimal media containing $(^{15}NH_4)_2SO_4$ as a sole nitrogen source. Purification of GST-fused proteins has been carried out according to previously published protocol (Kang, et al., 2003). All NMR samples

contained between 0.2 and 0.6 mM protein, 1 mM DTT and 150 mM NaCl in either 50 mM phosphate buffer, pH 6.5 or 50 mM TRIS buffer, pH 7.5.

**NMR spectroscopy**

NMR spectra were collected using Varian Inova 500 and 600 MHz spectrometers. NMR experiments have been typically collected at 30° C. Chemical shift assignment was based on a series of 3D experiments: $^{15}$N-separated HSQC-TOCSY (Zhang, et al., 1994), $^{15}$N-separated HSQC-NOESY with 150 ms mixing time (Zhang, et al., 1994) and HNHA (Vuister and Bax, 1993). Measurement of residual dipolar couplings was based on 2D IPAP-type $^{1}$H-$^{15}$N HSQC spectra (Ottiger, et al., 1998). Transformation of and analysis of NMR spectra was carried out in NMRPipe (Delaglio, et al., 1995) and Sparky (Goddard, T. D. & Kneller, J. M. University of California, San Francisco) programs.

**Relaxation and dynamics**

A standard set of experiments at 600 MHz was used to collect data for determination of $^{15}$N T1, T2 relaxation times and $^{15}$N{1H} NOE (Farrow, et al., 1994). Measurements have been carried out for 0.5 mM proteins in 50 mM Tris buffer, pH 6.5, 150 mM NaCl with addition of 1 mM DTT. Two sets of measurements for PDZ1, PDZ2 and PDZ12 have been carried out at 25 and 30°C. T1 relaxation was obtained from a series of experiments with 10, 80, 160, 240, 400, 650, 900 and 1400 ms time delays. For T2 measurements seven delays were used: 10, 30, 50, 90, 130, 170, 230 ms. $^{15}$N{$^{1}$H} NOE was obtained by recording one experiment with very little excitation of protons and a second experiment at saturating power with 3 s irradiation period (Farrow, et al., 1994).

Experimental correlation times were calculated from R2/R1 ratio using r2r1_tm program (Lee, et al., 1997). Prediction of global correlation times for monomeric and dimeric structures of PDZ tandem were obtained using the HYDRONMR program, assuming axially symmetric models (Garcia de la Torre, et al., 2000).

**Alignment in polyacrylamide gels**

Weak protein alignment was achieved in compressed copolymer polyacrylamide gels. Two types of positively charged copolymers have been used: 50% (3-acrylamidopropyl)-trimethylammonium chloride - 50% acrylamide (referred to hereafter as 50+M) and 75% (3-acrylamidopropyl)-trimethylammonium chloride - 25% acrylamide (75+M). Vertical gel compression has been achieved using Shigemi plunger (Sass, et al., 2000, Ishii, et al., 2001). The details of this method will be described elsewhere.

**Analysis of RDC**

Values of $^1D_{HN}$ were calculated from the difference between coupling constants measured in the absence and the presence of the gel. Determination of alignment tensor parameters ($D_a$ and R) and Euler angles ($\alpha$, $\beta$, $\gamma$) defining rotations of molecular coordinates about x, y and z axes, relative to principal axes frame of alignment tensor was carried out using the PALES program (Zweckstetter and Bax, 2000) and the crystal structure of PDZ tandem (PDB code 1N99). Compatibility between experimental and calculated RDC was evaluated based on quality factors Q calculated from the formula: $Q=rms(D^{calc}-D^{obs})/rms(D^{obs})$ (Cornilescu, et al., 1998).

**Analysis of domain orientation**

Initially, we carried out fitting of experimental $^1D_{HN}$ to crystal structures of individual PDZ domains. For calculation of quality factors we selected PDZ12$^A$ as the best representative of PDZ domain structures in solution. All subsequent rotations have been performed relative to the coordinate system from the original crystal structure with the origin located near the linker between PDZ1 and PDZ2. Euler angles determined for PDZ1 were used to rotate the tandem structure to the principal axis frame of PDZ1 alignment tensor. For the transformed coordinates we calculated Euler angles for PDZ2 defining the rotation necessary to match alignment of PDZ1. Applying the transformation to PDZ2 only we obtained tandem structure with reoriented domains that is consistent with RDC data. Evaluation of the error in determination of alignment tensor and domain orientation has been done using jack-knife procedure (Mosteller and Tukey, 1977) by performing 100 cycles of calculation with random elimination of 10 % of the data.

**Structure calculation**

The calculation of the syntenin tandem structure has been carried out using low temperature simulated annealing using rigid body structures of individual domains (Goto, et al., 2001) with addition of ambiguous interdomain distance restraints derived from $^1H^N/^{15}N$ chemical shift changes (Clore and Schwieters, 2003). In order to keep a rigid backbone conformation of the PDZ domains, we applied tight distance and dihedral angle restraints derived from the crystal structures. Separate sets of constraints for PDZ1 (residues 113-193) and PDZ2 (residues 198-270) were generated in the program Molmol (Koradi, et al., 1996). Distance constraints were created for all pairs of backbone N, $C^\alpha$

and C atoms separated by less than 6 Å and the dihedral angles were composed for $\phi$, $\psi$ and $\omega$ angles. Furthermore, we included a list of residual dipolar couplings and ambiguous distance restraints describing interdomain contacts. In order to transform information from chemical shift analysis into highly ambiguous distance restraints, we used a procedure similar to that proposed by Clore and Schwieters (Clore and Schwieters, 2003). The difference in chemical shifts between isolated PDZ domains and the tandem have been calculated using the following formula: $\Delta\delta=|\Delta\delta_{HN}|+0.11*|\Delta\delta_{N}|$. Highly ambiguous distance restraints were created between all amides from PDZ1 and PDZ2 with $\Delta\delta>0.08$ ppm. For the constructed distance we arbitrarily assigned lower and upper bounds of 2 and 5 Å, respectively.

All calculations were performed using CNS (Brunger, et al., 1998) with similar parameters to those described previously (Goto, et al., 2001). Torsion angle dynamics at 200 K for 15 ps and 3 fs time step was followed by a slow cooling stage in 5 K steps during a 150 ps simulation. Van der Waals scaling factor was ramped from 0.1 to 1 kcal/mol, while force constant for dipolar couplings was ramped from 0.05 to 0.5 kcal/mol $Hz^2$. Distances and dihedral angles were restrained with a force constant of 200 kcal/mol $A^2$ and 500 kcal/mol $rad^2$, respectively. At a final stage the structures were minimized with ten cycles of conjugate gradient minimization.

The initial model of the PDZ tandem oriented using RDCs has been used to calculate 100 structures. A family of ten structures with the best agreement between experimental and calculated RDCs has been selected for analysis. No distance violations larger than 0.2 Å or dihedral angle violations exceeding 5° were observed in the calculated structures. Separate calculations were performed for each set of RDCs.

## Acknowledgments

This work was supported by DOD grant DAMD17-01-1-0720.

## References

Bax, A. and Tjandra, N. (1997). High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J. Biomol. NMR *10*, 289-292.

Bax, A., Kontaxis, G. and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. Methods. Enzymol. *339*, 127-174.

Bax, A. (2003). Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein. Sci. *12*, 1-16.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Cryst.. *D54*, 905-921.

Cierpicki, T., Zhukov, I., Byrd, R.A. and Otlewski, J. (2002). Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. J. Magn. Reson. *157*, 178-180.

Clore, G.M., Starich, M.R. and Gronenborn, A.N. (1998). Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod- shape viruses. J. Am. Chem. Soc. *120*, 10571-10572.

Clore, G.M. and Schwieters, C.D. (2003). Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from $^{1}H/^{15}N$ chemical shift mapping and backbone $^{15}N$-$^{1}H$ residual dipolar couplings using conjoined rigid body/torsion angle dynamics. J. Am. Chem. Soc. *125*, 2902-2912.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J. Am. Chem. Soc. *120*, 6836-6837.

de Alba, E. and Tjandra, N. (2002). NMR dipolar couplings for the structure determination of biopolymers in solution. Prog. in Nucl. Magn. Reson. Spectroscopy *40*, 175-197.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277-293.

Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M. and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell *85*, 1067-1076.

Evenas, J., Tugarinov, V., Skrynnikov, N.R., Goto, N.K., Muhandiram, R. and Kay, L.E. (2001). Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy. J. Mol. Biol. *309*, 961-974.

Fan, J.S. and Zhang, M. (2002). Signaling complex organization by PDZ domain proteins. Neurosignals *11*, 315-321.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by $^{15}N$ NMR relaxation. Biochemistry *33*, 5984-6003.

Feng, W., Shi, Y., Li, M. and Zhang, M. (2003). Tandem PDZ repeats in glutamate receptor-interacting proteins have a novel mode of PDZ domain-mediated target binding. Nat. Struct. Biol. *10*, 972-978.

Fischer, M.W., Losonczi, J.A., Weaver, J.L. and Prestegard, J.H. (1999). Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. Biochemistry *38*, 9013-9022.

Garcia de la Torre, J., Huertas, M.L. and Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J. Magn. Reson. *147*, 138-146.

Geijsen, N., Uings, I.J., Pals, C., Armstrong, J., McKinnon, M., Raaijmakers, J.A., Lammers, J.W., Koenderman, L. and Coffer, P.J. (2001). Cytokine-specific transcriptional regulation through an IL-5Ralpha interacting protein. Science *293*, 1136-1138.

Goto, N.K., Skrynnikov, N.R., Dahlquist, F.W. and Kay, L.E. (2001). What is the average conformation of bacteriophage T4 lysozyme in solution? A domain orientation study using dipolar couplings measured by solution NMR. J. Mol. Biol. *308*, 745-764.

Grootjans, J.J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Durr, J. and David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. Proc. Natl. Acad. Sci. U S A *94*, 13683-13688.

Grootjans, J.J., Reekmans, G., Ceulemans, H. and David, G. (2000). Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin. J. Biol. Chem. *275*, 19933-19941.

Hillier, B.J., Christopherson, K.S., Prehoda, K.E., Bredt, D.S. and Lim, W.A. (1999). Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science *284*, 812-815.

Im, Y.J., Lee, J.H., Park, S.H., Park, S.J., Rho, S.H., Kang, G.B., Kim, E. and Eom, S.H. (2003). Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. J. Biol. Chem. *278*, 48099-48104.

Imamura, F., Maeda, S., Doi, T. and Fujiyoshi, Y. (2002). Ligand binding of the second PDZ domain regulates clustering of PSD-95 with the Kv1.4 potassium channel. J. Biol. Chem. *277*, 3640-3646.

Ishii, Y., Markus, M.A. and Tycko, R. (2001). Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel. J. Biomol. NMR *21*, 141-151.

Jelen, F., Oleksy, A., Smietana, K. and Otlewski, J. (2003). PDZ domains - common players in the cell signaling. Acta. Biochim. Pol. *50*, 985-1017.

Kang, B.S., Cooper, D.R., Devedjiev, Y., Derewenda, U. and Derewenda, Z.S. (2003). Molecular roots of degenerate specificity in syntenin's PDZ2 domain: reassessment of the PDZ recognition paradigm. Structure, *11*, 845-853.

Kang, B.S., Cooper, D.R., Jelen, F., Devedjiev, Y., Derewenda, U., Dauter, Z., Otlewski, J. and Derewenda, Z.S. (2003). PDZ tandem of human syntenin: crystal structure and functional properties. Structure, *11*, 459-468.

Kang, B.S., Devedjiev, Y., Derewenda, U. and Derewenda, Z.S. (2004). The PDZ2 Domain of Syntenin at Ultra-high Resolution: Bridging the Gap Between Macromolecular and Small Molecule Crystallography. J. Mol. Biol. *338*, 483-493.

Koradi, R., Billeter, M. and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51-55.

Koroll, M., Rathjen, F.G. and Volkmer, H. (2001). The neural cell recognition molecule neurofascin interacts with syntenin-1 but not with syntenin-2, both of which reveal self-associating activity. J. Biol. Chem. *276*, 10646-10654.

Lee, L.K., Rance, M., Chazin, W.J. and Palmer, A.G., 3rd (1997). Rotational diffusion anisotropy of proteins from simultaneous analysis of $^{15}N$ and $^{13}C$ alpha nuclear spin relaxation. J. Biomol. NMR *9*, 287-298.

Lin, D., Gish, G.D., Songyang, Z. and Pawson, T. (1999). The carboxyl terminus of B class ephrins constitutes a PDZ domain binding motif. J. Biol. Chem. *274*, 3726-3733.

Long, J.F., Tochio, H., Wang, P., Fan, J.S., Sala, C., Niethammer, M., Sheng, M. and Zhang, M. (2003). Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. J. Mol. Biol. *327*, 203-214.

Losonczi, J.A., Andrec, M., Fischer, M.W. and Prestegard, J.H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. J. Magn. Reson. *138*, 334-342.

Meier, S., Haussinger, D. and Grzesiek, S. (2002). Charged acrylamide copolymer gels as media for weak alignment. J. Biomol. NMR *24*, 351-356.

Mosteller, F. and Tukey, J. (1977). *Data analysis and regression: a second course in statistics*. Addison-Wesley Publishing Co., Don Mills, Ontario.

Nourry, C., Grant, S.G. and Borg, J.P. (2003). PDZ domain proteins: plug and play! Sci. STKE *2003*, RE7.

Ottiger, M., Delaglio, F. and Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J. Magn. Reson. *131*, 373-378.

Ottiger, M. and Bax, A. (1999). Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. J. Biomol. NMR *13*, 187-191.

Prestegard, J.H., al-Hashimi, H.M. and Tolman, J.R. (2000). NMR structures of biomolecules using field oriented media and residual dipolar couplings. Q. Rev. Biophys. *33*, 371-424.

Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J. Biomol. NMR *18*, 303-309.

Ulmer, T.S., Ramirez, B.E., Delaglio, F. and Bax, A. (2003). Evaluation of backbone proton positions and dynamics in a small protein by liquid crystall NMR spectroscopy. J. Am. Chem. Soc. *125*, 9179-9191.

Vuister, G.W. and Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond $J_{HNH\alpha}$ coupling constants in $^{15}N$-enriched proteins. J. Am. Chem. Soc. *115*, 7772-7777.

Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994). Backbone $^{1}H$ and $^{15}N$ resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J. Biomol. NMR *4*, 845-858.

Zimmermann, P., Tomatis, D., Rosas, M., Grootjans, J., Leenaerts, I., Degeest, G., Reekmans, G., Coomans, C. and David, G. (2001). Characterization of syntenin, a syndecan-binding PDZ protein, as a component of cell adhesion sites and microfilaments. Mol. Biol. Cell *12*, 339-350.

Zweckstetter, M. and Bax, A. (2000). Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J. Am. Chem. Soc. *122*, 3791-3792.

## Figure Legends

**Figure1.** Crystal structure of the syntenin PDZ tandem dimer with residues colored according to the extent of differences in chemical shifts between isolated PDZ domains and in the tandem. Chemical shift differences were calculated according to the formula $\Delta\delta=|\Delta\delta_{HN}|+0.11*|\Delta\delta_{N}|$. Red (PDZ1) and blue (PDZ2) correspond to residues with the largest chemical shift changes ($\Delta\delta>0.16$ ppm) while yellow (PDZ1) and cyan (PDZ2) indicate smaller differences ($0.16>\Delta\delta>0.08$ ppm).

**Figure 2.** Solution structure of the syntenin PDZ tandem. A) superposition of 17 lowest energy structures calculated using two sets of residual dipolar couplings: 50+M (magenta, 10 structures) and 75+M (green, 7 structures). B) difference in PDZ2 orientation between the crystal structure of the PDZ tandem (green) and the structure in solution (blue). PDZ2 is rotated by -5°, 3° and -23° about x, y and z axes, respectively, in the indicated coordinate system. All structures were superimposed on the PDZ1 domain (backbone residues 113-193).

**Figure 3.** The $^{1}H$-$^{15}N$ HSQC spectrum of 32 kDa full-length syntenin measured at 30°C for 0.2 mM protein in 50 mM Tris buffer, pH 7.5 and 150 mM NaCl. The yellow field covers the severely overlapped region containing unstructured N-terminal residues. The upfiled shifted signal at 6.34 ppm (green) is present only in the spectrum of full-length syntenin.

**Figure 4.** Comparison of $^{1}H$-$^{15}N$ HSQC spectra of PDZ12 (black) with N-PDZ12 (red), PDZ12-C (green) and full length syntenin (blue). All spectra are recorded at 30°C under identical conditions (50 mM Tris buffer, 150 mM NaCl, pH 7.5).

## Tables

**Table I.** Alignment tensor parameters calculated for PDZ1, PDZ2 and PDZ12 based on RDC measured for PDZ domain tandem weakly aligned in 75+M gel. $D_a$ is magnitude of alignment tensor normalized to $^1D_{HN}$, R is rhombicity, Q – quality factor and $\alpha$, $\beta$, $\gamma$ are Euler angles defining rotation about x, y and z axes, respectively. a) Euler angles calculated for PDZ2 in the tandem structure transformed to principal axis frame of PDZ1; b) parameters of alignment tensor for PDZ12 upon reorientation of PDZ2 in order to match the alignment frame of PDZ1 (values of Euler angles are identical to that of PDZ1 and not shown).

| | PDZ1 | PDZ2 | PDZ12 | PDZ2[a] | PDZ12[b] |
|---|---|---|---|---|---|
| $D_a$ | 6.20 | 6.67 | 6.49 | 6.67 | 6.48 |
| R | 0.58 | 0.46 | 0.44 | 0.46 | 0.61 |
| Q | 22.3 | 22.1 | 27.7 | 22.1 | 22.7 |
| $\alpha$ | -26.0±0.4 | -32.0±0.6 | -29.5±0.4 | -4.6±0.3 | - |
| $\beta$ | 18.0±0.4 | 14.4±0.6 | 17.8±0.4 | 3.4±0.7 | - |
| $\gamma$ | -42.2±0.7 | -66.6±0.8 | -55.5±0.9 | -22.7±0.8 | - |

**Figure 1.**

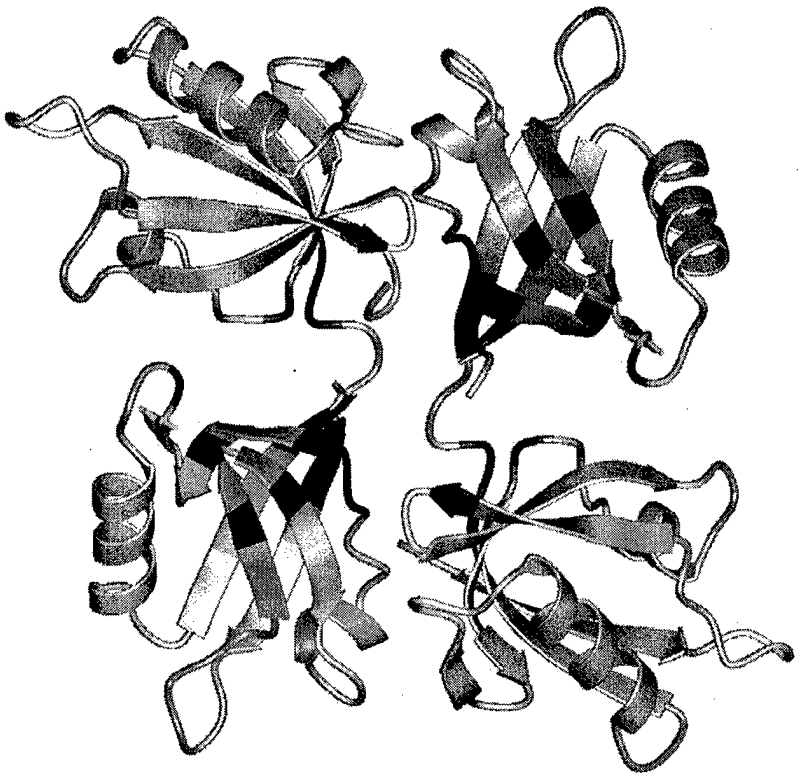


**Figure 2.**

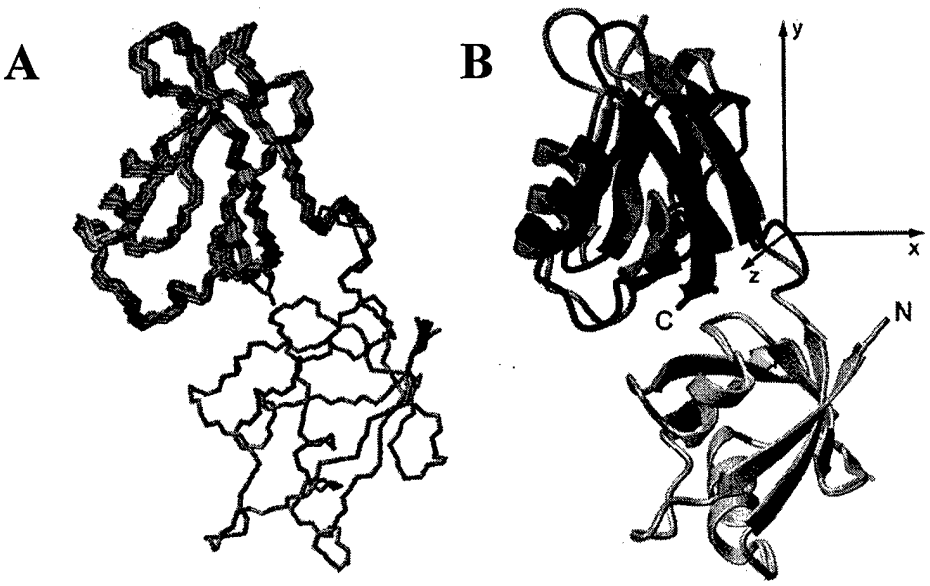


**Figure 3.**

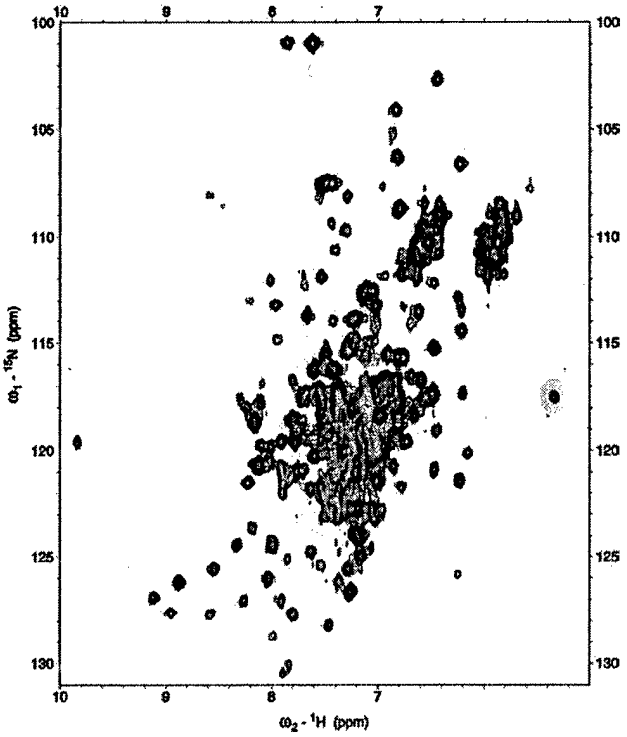
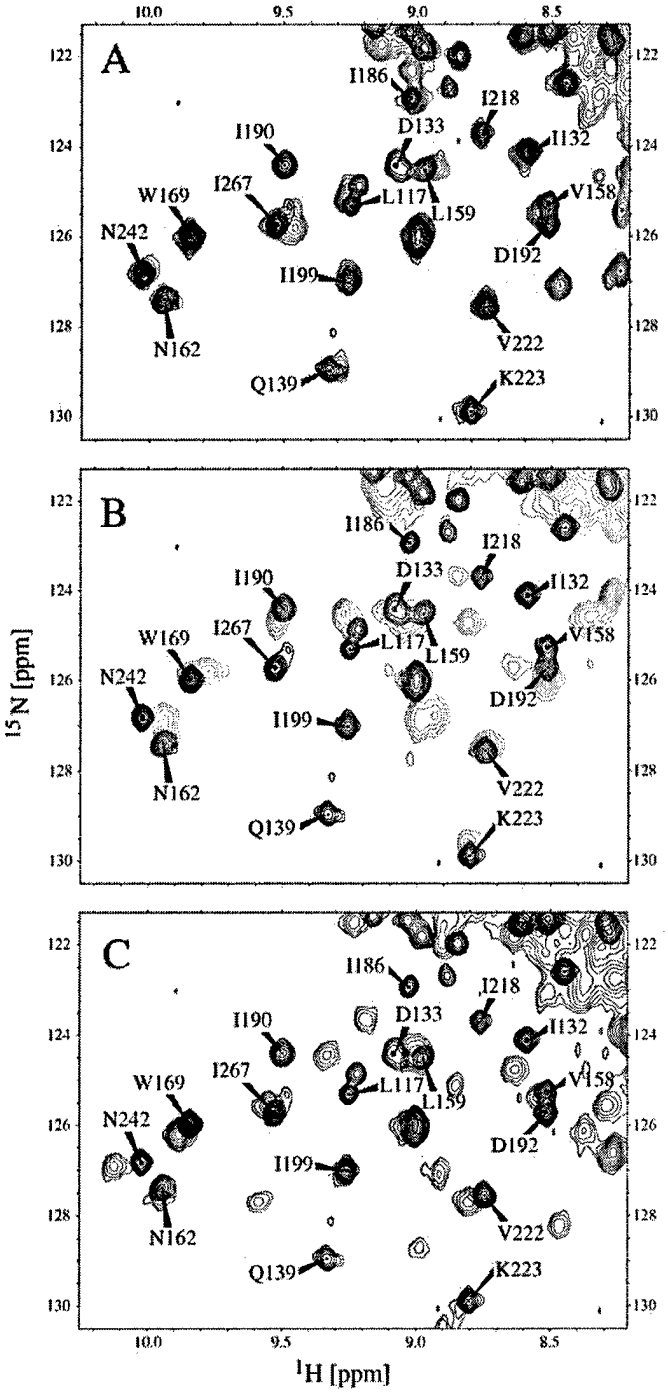


**Figure 4.**


A
B
C
I186
I218
D133
I190
I132
I267
W169
L117
V158
L159
N242
D192
I199
N162
V222
Q139
K223
10.0
9.5
9.0
8.5
122
124
126
128
130
$^{15}$N [ppm]
$^{1}$H [ppm]